

A Rate-Splitting Approach to Fading Channels with Imperfect Channel-State Information

Adriano Pastore, *Student Member, IEEE*, Tobias Koch, *Member, IEEE*,
and Javier Rodríguez Fonollosa, *Senior Member, IEEE*

Abstract

As shown by Médard (“The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. Inf. Theory*, May 2000), the capacity of fading channels with imperfect channel-state information (CSI) can be lower-bounded by assuming a Gaussian channel input X with power P and by upper-bounding the conditional entropy $h(X|Y, \hat{H})$, conditioned on the channel output Y and the CSI \hat{H} , by the entropy of a Gaussian random variable with variance equal to the linear minimum mean-square error in estimating X from (Y, \hat{H}) . We demonstrate that, using a rate-splitting approach, this lower bound can be sharpened: by expressing the Gaussian input X as the sum of two independent Gaussian variables X_1 and X_2 and by applying Médard’s lower bound first to bound the mutual information between X_1 and Y while treating X_2 as noise, and by applying the lower bound then to bound the mutual information between X_2 and Y while assuming X_1 to be known, we obtain a lower bound on the capacity that is strictly larger than Médard’s lower bound. We then generalize this approach to an arbitrary number L of layers, where X is expressed as the sum of L independent Gaussian random variables of respective variances P_ℓ , $\ell = 1, \dots, L$ summing up to P . Among all such rate-splitting bounds, we determine the supremum over power allocations P_ℓ and total number of layers L . This supremum is achieved for $L \rightarrow \infty$ and gives rise to an analytically expressible lower bound on the Gaussian-input mutual information. For Gaussian fading, this novel bound is shown to be asymptotically tight at high signal-to-noise ratio (SNR), provided that the variance of the channel estimation error $H - \hat{H}$ tends to zero as the SNR tends to infinity.

A. Pastore and J. R. Fonollosa have been supported by the Ministerio de Economía y Competitividad of Spain (TEC2010-19171 and CONSOLIDER-INGENIO 2010 CSD2008-00010 COMONSENS) and Generalitat de Catalunya (2009SGR-1236). T. Koch has been supported by the Ministerio de Economía y Competitividad of Spain (TEC2009-14504-C02-01 DEIPRO and CONSOLIDER-INGENIO 2010 CSD2008-00010 COMONSENS). The material in this paper was presented in part at the IEEE 27th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, November 14–17, 2012.

A. Pastore and J. R. Fonollosa are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Jordi Girona 1-3, Building D5, 08034 Barcelona, Spain (email: {adriano.pastore,javier.fonollosa}@upc.edu).

T. Koch is with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911 Leganés, Spain (email: koch@tsc.uc3m.es).

I. INTRODUCTION AND CHANNEL MODEL

We consider a single-antenna memoryless fading channel with imperfect channel-state information (CSI), whose time- k channel output $Y[k]$ corresponding to a time- k channel input $X[k] = x \in \mathbb{C}$ (where \mathbb{C} denotes the set of complex numbers) is given by

$$Y[k] = (\hat{H}[k] + \tilde{H}[k])x + Z[k], \quad k \in \mathbb{Z} \quad (1)$$

(with \mathbb{Z} denoting the set of integers). Here, the noise $\{Z[k]\}_{k \in \mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.), zero-mean, circularly-symmetric, complex Gaussian random variables with variance $\mathbb{E}[|Z[k]|^2] = N_0$. The fading pair $\{(\hat{H}[k], \tilde{H}[k])\}_{k \in \mathbb{Z}}$ is an arbitrary sequence of i.i.d. complex-valued random variables whose means and variances satisfy the following conditions:

- $\hat{H}[k]$ has mean μ and variance \hat{V} ;
- conditioned on $\hat{H}[k] = \hat{h}$, the random variable $\tilde{H}[k]$ has zero mean and variance $\tilde{V}(\hat{h})$, i.e.,

$$\mathbb{E}[\tilde{H}[k] \mid \hat{H}[k] = \hat{h}] = 0 \quad (2a)$$

$$\mathbb{E}[|\tilde{H}[k]|^2 \mid \hat{H}[k] = \hat{h}] \triangleq \tilde{V}(\hat{h}). \quad (2b)$$

We assume that the joint sequence $\{(\hat{H}[k], \tilde{H}[k])\}_{k \in \mathbb{Z}}$, the noise sequence $\{Z[k]\}_{k \in \mathbb{Z}}$ and the input sequence $\{X[k]\}_{k \in \mathbb{Z}}$ are all three mutually independent. We further assume that the receiver is cognizant of the realization of $\{\hat{H}[k]\}_{k \in \mathbb{Z}}$, but the transmitter is only cognizant of its distribution. We finally assume that both the transmitter and receiver are cognizant of the distributions of $\{\tilde{H}[k]\}_{k \in \mathbb{Z}}$ and $\{Z[k]\}_{k \in \mathbb{Z}}$ but not of their realizations. The $\hat{H}[k]$ can be viewed as an estimate of the channel fading coefficient

$$H[k] \triangleq \hat{H}[k] + \tilde{H}[k]. \quad (3)$$

Accordingly, $\tilde{H}[k]$ can be viewed as the channel estimation error. From this perspective, the condition (2a) is, for example, satisfied when $\hat{H}[k]$ is the minimum mean-square error (MMSE) estimate of $H[k]$ from some receiver side information. When $\tilde{H}[k] = 0$ almost surely, we shall say that the receiver has perfect CSI.

The capacity of the above channel (1) under the average-power constraint P on the channel inputs is given by [1]

$$C(P) = \sup I(X; Y | \hat{H}) \quad (4)$$

where the supremum is over all distributions of X satisfying $\mathbb{E}[|X|^2] \leq P$. Here and throughout the paper we omit the time indices k wherever they are immaterial. Since (4) is difficult to evaluate, even if \hat{H} and \tilde{H} are Gaussian, it is common to assess $C(P)$ using upper and lower bounds. A widely-used lower bound on $C(P)$ is due to Médard [2]:

$$C(P) \geq \mathbb{E} \left[\log \left(1 + \frac{|\hat{H}|^2 P}{\tilde{V}(\hat{H})P + N_0} \right) \right] \triangleq R_M(P). \quad (5)$$

This lower bound follows from (4) by choosing X to be zero-mean, variance- P Gaussian and by upper-bounding the differential entropy of X conditioned on Y and \hat{H} as

$$\begin{aligned} h(X|Y, \hat{H}) &= h(X - \alpha Y|Y, \hat{H}) \\ &\leq h(X - \alpha Y|\hat{H}) \\ &\leq \mathbb{E} \left[\log \left(\pi e \mathbb{E}[|X - \alpha Y|^2 | \hat{H}] \right) \right] \end{aligned} \quad (6)$$

for any $\alpha \in \mathbb{C}$. Here the first inequality follows because conditioning cannot increase entropy, and the subsequent inequality follows because the Gaussian distribution maximizes differential entropy for a given second moment [3, Th. 9.6.5]. By expressing the mutual information $I(X; Y|\hat{H})$ as

$$I(X; Y|\hat{H}) = h(X) - h(X|Y, \hat{H}) \quad (7)$$

and by choosing α so that αY is the linear MMSE estimate of X , the lower bound (5) follows.

When the receiver has perfect CSI so that $\mathbb{E}[\tilde{V}(\hat{H})] = 0$, the lower bound $R_M(P)$ is equal to the channel capacity

$$C_{\text{coh}}(P) = \mathbb{E} \left[\log \left(1 + \frac{|H|^2 P}{N_0} \right) \right]. \quad (8)$$

Consequently, for perfect CSI the lower bound (5) is tight.

In contrast, when the receiver has imperfect CSI and $\tilde{V}(\hat{H})$ and \hat{H} do not depend on P , the lower bound (5) is loose. In fact, in this case the lower bound $R_M(P)$ is bounded in P , whereas the capacity $C(P)$ is known to be unbounded. For instance, if \tilde{H} is of finite differential entropy, then the capacity has a double-logarithmic growth in P [4].¹

This boundedness of $R_M(P)$ is not due to the inequalities in (6) being loose, but is a consequence of choosing a Gaussian channel input. Indeed, if \tilde{H} is of finite differential entropy, then a Gaussian input X_G achieves [5, Proposition 6.3.1], [4, Lemma 4.5]

$$\overline{\lim}_{P \rightarrow \infty} I(X_G; Y|\hat{H}) \leq \gamma + \log(\pi e \mathbb{E}[|\hat{H} + \tilde{H}|^2]) - h(\tilde{H}) \quad (9)$$

where $\gamma \approx 0.577$ denotes Euler's constant and where $\overline{\lim}$ denotes the *limit superior*. Nevertheless, even if we restrict ourselves to Gaussian inputs, the lower bound

$$I(X_G; Y|\hat{H}) \geq R_M(P) \quad (10)$$

is not tight. As we shall see, by using a rate-splitting (or successive-decoding) approach, this lower bound (10) can be sharpened: we show in Section II that, by expressing the Gaussian input X_G as the sum of two independent Gaussian random variables X_1 and X_2 , and by first applying the bounding technique sketched in (6)–(7) to $I(X_1; Y|\hat{H})$ (thus treating $\tilde{H}X_2$ as noise) and then using the same bounding technique to lower-bound $I(X_2; Y|\hat{H}, X_1)$, we obtain

¹This result can be generalized to show that if $\mathbb{E}[\log |\hat{H} + \tilde{H}|^2] > -\infty$ holds, then the capacity grows at least double-logarithmically with P .

a lower bound on the Gaussian-input mutual information (and thus also on the capacity) that is strictly larger than the conventional bound $R_M(P)$.

In Section III, we expand this approach by expressing X as the sum of $L \geq 2$ independent Gaussian random variables X_ℓ , $\ell = 1, \dots, L$ and by applying the bounding technique from (6)–(7) first to $I(X_1; Y|\hat{H})$, then to $I(X_2; Y|\hat{H}, X_1)$, and so on. We show that the so obtained lower bound is strictly increasing in L (provided that we optimize the sum of bounds over the powers $P_\ell = \mathbb{E}[|X_\ell|^2]$, $\ell = 1, \dots, L$), and we determine its limit as L tends to infinity. The so-obtained lower bound permits an analytic expression. In the remainder of this paper, we shall refer to ℓ as a *layer* and to L as the *number of layers*.

In Section IV, we show that when, conditioned on \hat{H} , the estimation error \tilde{H} is Gaussian, and when its average variance (averaged over \hat{H}) tends to zero as the SNR tends to infinity, the new lower bound tends to the Gaussian-input mutual information $I(X_G; Y|\hat{H})$ as the SNR tends to infinity. For non-Gaussian fading, we show that, at high SNR, the difference between $I(X_G; Y|\hat{H})$ and our lower bound is upper-bounded by the difference of the logarithms of the variance of \tilde{H} and of its entropy power.

The rest of the paper is organized as follows. In Section V we discuss the connection of our results with similar results obtained in the mismatched-decoding literature. In Sections VI and VII we provide the proofs of the main results. And in Section VIII we conclude the paper with a summary and discussion.

II. RATE-SPLITTING WITH TWO LAYERS

For future reference, we state Médard's lower bound (5) in a slightly more general form in the following proposition.

Proposition 1 (Médard [2]). *Let S be a zero-mean, circularly-symmetric, complex Gaussian random variable of variance P . Let A and B be complex-valued random variables of finite second moments, and let C be an arbitrary random variable. Assume that S is independent of (A, C) , and that, conditioned on (A, C) , the variables S and B are uncorrelated. Then*

$$I(S; AS + B|A, C) \geq \mathbb{E} \left[\log \left(1 + \frac{|A|^2 P}{V_B(A, C)} \right) \right] \quad (11)$$

where $V_B(a, c)$ denotes the conditional variance of B conditioned on $(A, C) = (a, c)$.

Proof: See Appendix A. ■

Using Proposition 1, we show that, for imperfect CSI and $\mathbb{E}[|\hat{H}|^2] > 0$, rate splitting with two layers strictly improves the lower bound (10). Indeed, let X_1 and X_2 be independent, zero-mean, circularly-symmetric, complex Gaussian random variables with respective variances P_1 and P_2 (satisfying $P_1 + P_2 = P$) such that $X = X_1 + X_2$. By the chain rule for mutual information, we obtain

$$\begin{aligned} I(X; Y|\hat{H}) &= I(X_1, X_2; Y|\hat{H}) \\ &= I(X_1; Y|\hat{H}) + I(X_2; Y|\hat{H}, X_1). \end{aligned} \quad (12)$$

By replacing the variables A , B , C and S in Proposition 1 with

$$A \leftarrow \hat{H}, \quad B \leftarrow \hat{H}X_2 + \tilde{H}X + Z, \quad C \leftarrow 0, \quad S \leftarrow X_1$$

it follows that the first term on the right-hand side (RHS) of (12) is lower-bounded as

$$I(X_1; Y | \hat{H}) \geq \mathbb{E} \left[\log \left(1 + \frac{|\hat{H}|^2 P_1}{(|\hat{H}|^2 + \tilde{V}(\hat{H}))P_2 + \tilde{V}(\hat{H})P_1 + N_0} \right) \right] \triangleq R_1(P_1, P_2). \quad (13)$$

Similarly, by replacing A , B , C , S in Proposition 1 with

$$A \leftarrow \hat{H}, \quad B \leftarrow \hat{H}X_1 + \tilde{H}X + Z, \quad C \leftarrow X_1, \quad S \leftarrow X_2$$

we obtain for the second term on the RHS of (12)

$$I(X_2; Y | \hat{H}, X_1) \geq \mathbb{E} \left[\log \left(1 + \frac{|\hat{H}|^2 P_2}{\tilde{V}(\hat{H})(|X_1|^2 + P_2) + N_0} \right) \right] \triangleq R_2(P_1, P_2). \quad (14)$$

Noting that, for every $\alpha > 0$, the function $x \mapsto \log(1 + \alpha/x)$ is strictly convex in $x \geq 0$, it follows from Jensen's inequality that the RHS of (14) is lower-bounded as

$$\mathbb{E} \left[\log \left(1 + \frac{|\hat{H}|^2 P_2}{\tilde{V}(\hat{H})(|X_1|^2 + P_2) + N_0} \right) \right] \geq \mathbb{E} \left[\log \left(1 + \frac{|\hat{H}|^2 P_2}{\tilde{V}(\hat{H})(P_1 + P_2) + N_0} \right) \right] \quad (15)$$

with the inequality being strict except in the trivial cases where $P_1 = 0$, $P_2 = 0$, or if, with probability one, at least one of $|\hat{H}|$ and $\tilde{V}(\hat{H})$ is zero.² Thus, combining (12)–(15), we obtain

$$R_1(P_1, P_2) + R_2(P_1, P_2) \geq \mathbb{E} \left[\log \left(1 + \frac{|\hat{H}|^2 P}{\tilde{V}(\hat{H})P + N_0} \right) \right] \quad (16)$$

demonstrating that, when the receiver has imperfect CSI, rate splitting with two layers strictly improves the capacity and mutual information lower bound (5) (except in trivial cases).

Figure 1 compares the two-layer bound $R_1(P_1, P_2) + R_2(P_1, P_2)$ with $R_M(P)$ (dashed line) as a function of P_1/P , for \hat{H} and \tilde{H} being circularly-symmetric Gaussian with parameters $\mu = 0$, $\hat{V} = \frac{1}{2}$, $\tilde{V}(\hat{h}) = \frac{1}{2}$ for $\hat{h} \in \mathbb{C}$, $P = 10$, and $N_0 = 1$. The figure confirms our above observation that, when the receiver has imperfect CSI and $P_1 > 0$ and $P_2 > 0$, rate splitting with two layers outperforms $R_M(P)$ (5). In this example, the optimal power allocation is approximately at $P_1 \approx 0.78P$ and $P_2 \approx 0.22P$.

One might wonder whether extending our approach to more than two layers can further improve the lower bound. As we shall see in the following section, it does. In fact, for every power P we show that, once that the power is optimally allocated across layers, the rate-splitting lower bound is strictly increasing in the number of layers.

²We may write this as $\Pr\{\hat{H} \cdot \tilde{V}(\hat{H}) = 0\} = 1$. For example, this occurs when the receiver has perfect CSI, in which case $\tilde{V}(\hat{H}) = 0$ almost surely.

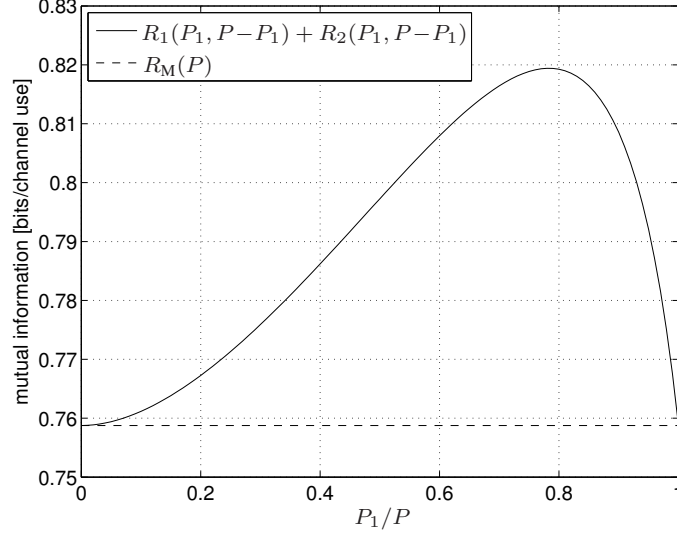


Fig. 1. Comparison of the 2-layer lower bound $R_1(P_1, P - P_1) + R_2(P_1, P - P_1)$ (continuous line) with Médard's lower bound $R_M(P)$ (dashed line) as a function of the power fraction P_1/P assigned to the first layer.

III. RATE-SPLITTING WITH L LAYERS

Let X_1, \dots, X_L be independent, zero-mean, circularly-symmetric, complex Gaussian random variables with respective variances P_1, \dots, P_L satisfying

$$P = \sum_{\ell=1}^L P_\ell \quad (17)$$

such that

$$X = \sum_{\ell=1}^L X_\ell.$$

Let the cumulative power Q_k be given by

$$Q_k = \sum_{\ell=1}^k P_\ell. \quad (18)$$

We denote the collection of cumulative powers as

$$\mathbf{Q} = (Q_1, \dots, Q_L) \quad (19)$$

and refer to it as an L -layering.

It follows from the chain rule for mutual information that

$$I(X^L; Y | \hat{H}) = \sum_{\ell=1}^L I(X_\ell; Y | X^{\ell-1}, \hat{H}) \quad (20)$$

where we use the shorthand A^N to denote the sequence A_1, \dots, A_N , and A^0 denotes the empty sequence. Applying Proposition 1 by replacing A, B, C, S with the respective

$$A \leftarrow \hat{H}, \quad B \leftarrow \hat{H} \sum_{\ell' \neq \ell} X_{\ell'} + \tilde{H}X + Z, \quad C \leftarrow X^{\ell-1}, \quad S \leftarrow X_\ell,$$

we can lower-bound the ℓ -th summand on the RHS of (20) as

$$\begin{aligned} I(X_\ell; Y | X^{\ell-1}, \hat{H}) &\geq \mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{Q}}(X^{\ell-1}, \hat{H})) \right] \\ &\triangleq R_\ell[\mathbf{Q}] \end{aligned} \quad (21)$$

where

$$\Gamma_{\ell, \mathbf{Q}}(X^{\ell-1}, \hat{H}) \triangleq \frac{|\hat{H}|^2 P_\ell}{\tilde{V}(\hat{H}) \left| \sum_{i < \ell} X_i \right|^2 + \tilde{V}(\hat{H}) P_\ell + (|\hat{H}|^2 + \tilde{V}(\hat{H})) \sum_{i > \ell} P_i + N_0} \quad (22)$$

and where the last line in (21) should be viewed as the definition of $R_\ell[\mathbf{Q}]$. Defining

$$R[\mathbf{Q}] \triangleq R_1[\mathbf{Q}] + \dots + R_L[\mathbf{Q}] \quad (23)$$

we obtain from (20) and (21) the lower bound

$$I(X; Y | \hat{H}) = I(X^L; Y | \hat{H}) \geq R[\mathbf{Q}]. \quad (24)$$

Note that $Q_{\ell-1} = Q_\ell$ implies $P_\ell = 0$, which in turn implies $R_\ell[\mathbf{Q}] = 0$. Without loss of optimality, we can therefore restrict ourselves to L -layerings satisfying

$$0 < Q_1 < \dots < Q_L = P. \quad (25)$$

We shall denote the set of such L -layerings by $\mathcal{Q}(P, L)$.

Let $R^*(P, L)$ denote the lower $R[\mathbf{Q}]$ optimized over all $\mathbf{Q} \in \mathcal{Q}(P, L)$, i.e.,

$$R^*(P, L) \triangleq \sup_{\mathbf{Q} \in \mathcal{Q}(P, L)} R[\mathbf{Q}]. \quad (26)$$

In the following, we show that $R^*(P, L)$ is monotonically increasing in L . To this end, we need the following lemma.

Lemma 2. *Let $L' > L$, and let the L -layering $\mathbf{Q} \in \mathcal{Q}(P, L)$ and the L' -layering $\mathbf{Q}' \in \mathcal{Q}(P, L')$ satisfy*

$$\{Q_1, \dots, Q_L\} \subset \{Q'_1, \dots, Q'_{L'}\}. \quad (27)$$

Then

$$R[\mathbf{Q}] \leq R[\mathbf{Q}'] \quad (28)$$

with equality if, and only if, $\Pr\{\hat{H} \cdot \tilde{V}(\hat{H}) = 0\} = 1$.

Proof: See Appendix B. ■

Theorem 3. *Assume that $\Pr\{\hat{H} \cdot \tilde{V}(\hat{H}) = 0\} < 1$. Then, $R^*(P, L)$ is monotonically nondecreasing in L .*

Proof: For every L -layering $\mathbf{Q} \in \mathcal{Q}(P, L)$, we can construct an $(L+1)$ -layering $\mathbf{Q}' \in \mathcal{Q}(P, L+1)$ satisfying $\mathbf{Q} \subset \mathbf{Q}'$ by adding $(Q_1 + Q_2)/2$ to \mathbf{Q} . Together with Lemma 2, this implies that for every $\mathbf{Q} \in \mathcal{Q}(P, L)$ there exists a $\mathbf{Q}' \in \mathcal{Q}(P, L+1)$ such that $R[\mathbf{Q}] < R[\mathbf{Q}']$, from which the theorem follows upon maximizing both sides of the inequality over all layerings $\mathbf{Q} \in \mathcal{Q}(P, L)$ and $\mathbf{Q}' \in \mathcal{Q}(P, L+1)$, respectively. ■

It follows from Theorem 3 that the best lower bound, optimized over all layerings of fixed sum-power P

$$R^*(P) \triangleq \sup_{L \in \mathbb{N}} \sup_{\mathbf{Q} \in \mathcal{Q}(P, L)} R[\mathbf{Q}] = \sup_{L \in \mathbb{N}} R^*(P, L) \quad (29)$$

(where \mathbb{N} denotes the positive integers) is approached by letting the number of layers L tend to infinity. An explicit expression for $R^*(P)$ is provided by the following theorem.

Theorem 4. *For a given input power P , the supremum of all rate-splitting lower bounds $R[\mathbf{Q}]$ over $\mathbf{Q} \in \mathcal{Q}(P, L)$ and $L \in \mathbb{N}$, is given by*

$$\begin{aligned} R^*(P) &= \lim_{L \rightarrow \infty} R^*(P, L) \\ &= \mathbb{E} \left[\frac{|\hat{H}|^2}{|\hat{H}|^2 + \tilde{V}(\hat{H}) + \frac{N_0}{P}} \Theta \left(\frac{\tilde{V}(\hat{H})(W - 1) - |\hat{H}|^2}{|\hat{H}|^2 + \tilde{V}(\hat{H}) + \frac{N_0}{P}} \right) \right] \end{aligned} \quad (30)$$

where

$$\Theta(x) \triangleq \begin{cases} \frac{1}{x} \log(1 + x), & \text{if } -1 < x < 0 \text{ or } x > 0 \\ 1, & \text{if } x = 0 \end{cases} \quad (31)$$

and where W is independent of \hat{H} and exponentially distributed with mean 1.

Proof: The proof of Theorem 4 is given in Section VI. ■

Remark 1. *The proof of Theorem 4 hinges on the observation that the supremum $R^*(P)$ is approached by a uniform layering*

$$\mathbf{U}(P, L) \triangleq \left(\frac{P}{L}, 2\frac{P}{L}, \dots, (L-1)\frac{P}{L}, P \right) \quad (32)$$

when the number of layers L is taken to infinity. While this layering was chosen for mathematical convenience, any other layering would also do, provided that some regularity conditions are met. For example, one can show that for any Lipschitz-continuous monotonic bijection $F: [0, P] \rightarrow [0, P]$, we have

$$\lim_{L \rightarrow \infty} R[F(\mathbf{U}(P, L))] = \lim_{L \rightarrow \infty} R[\mathbf{U}(P, L)] = R^*(P) \quad (33)$$

where $F(\mathbf{U}(P, L)) = \{F(P/L), F(2P/L), \dots, F(P)\}$.

To assess the tightness of the derived lower bounds, we consider two upper bounds on the mutual information for Gaussian inputs. The first upper bound is the capacity when the receiver has perfect CSI [cf. (8)] and follows by noting that improving the CSI at the receiver does not reduce mutual information:

$$I(X_G; Y | \hat{H}) \leq \mathbb{E} \left[\log \left(1 + \frac{|H|^2 P}{N_0} \right) \right] \triangleq C_{\text{coh}}(P). \quad (34)$$

The second upper bound is given by

$$\begin{aligned} I(X_G; Y | \hat{H}) &\leq R_M(P) + \mathbb{E} \left[\log \left(\frac{\tilde{V}(\hat{H})P + N_0}{\tilde{\Phi}(\hat{H})PW + N_0} \right) \right] \\ &\triangleq I_{\text{upper}}(P) \end{aligned} \quad (35)$$

where W is an exponentially distributed random variable of mean 1, and where $\tilde{\Phi}(\hat{h})$ denotes the conditional entropy power of \tilde{H} , conditioned on $\hat{H} = \hat{h}$:³

$$\tilde{\Phi}(\hat{h}) = \begin{cases} \frac{1}{\pi e} e^{h(\tilde{H}|\hat{H}=\hat{h})}, & \text{if } h(\tilde{H}|\hat{H}=\hat{h}) > -\infty \\ 0, & \text{otherwise.} \end{cases} \quad (36)$$

This upper bound follows from expanding the mutual information as $h(Y|\hat{H}) - h(Y|X_G, \hat{H})$, upper-bounding $h(Y|\hat{H})$ by the entropy of a Gaussian variable of same variance, and lower-bounding $h(Y|X_G, \hat{H})$ using the entropy-power inequality [6, Theorem 6].

The upper bound (35) was previously used, e.g., in [7, Equation (42)], [8, Lemma 2] for Gaussian fading, in which case the entropy-power inequality is tight and the entropy power equals the conditional variance, i.e., $\tilde{\Phi}(\hat{h}) = \tilde{V}(\hat{h})$ for $\hat{h} \in \mathbb{C}$.

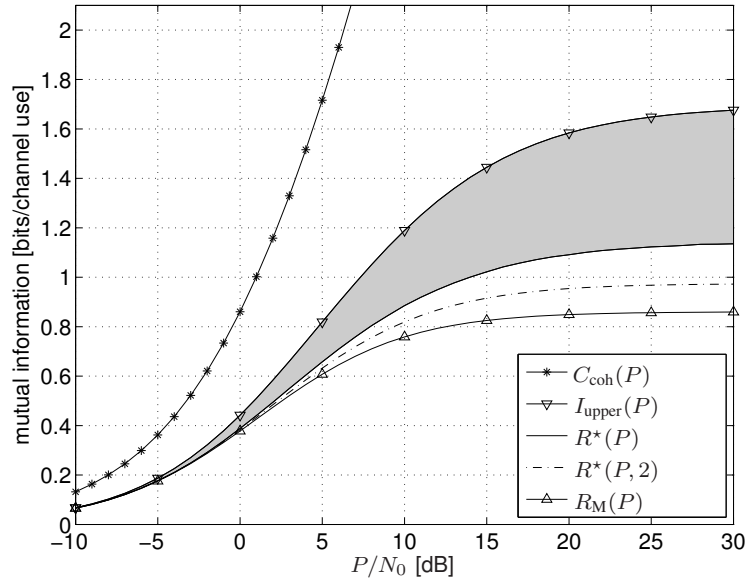


Fig. 2. Comparison between several bounds on the capacity and the Gaussian-input mutual information.

In Figure 2, several bounds on the mutual information $I(X_G; Y|\hat{H})$ for Gaussian inputs are plotted against the SNR on a range from -10 dB to 30 dB. From top to bottom, we have the coherent capacity (34); the upper bound (35); the supremum $R^*(P)$ of all rate-splitting bounds (Theorem 4); the two-layer rate-splitting bound with optimized power allocation $R^*(P, 2)$; and Médard's lower bound $R_M(P)$. The grey-shaded area indicates the region in which the curve of the exact Gaussian-input mutual information $I(X_G; Y|\hat{H})$ is located. For this simulation, we have chosen \hat{H} and \tilde{H} to be independent and complex circularly-symmetric Gaussian with parameters $\mu = 0$,

³We define $h(\tilde{H}|\hat{H}=\hat{h}) = -\infty$ if the conditional distribution of \tilde{H} , conditioned on $\hat{H} = \hat{h}$, is not absolutely continuous with respect to the Lebesgue measure.

$\hat{V} = \frac{1}{2}$, and $\tilde{V}(\hat{h}) = \frac{1}{2}$, $\hat{h} \in \mathbb{C}$. Observe that the proposed rate-splitting approach achieves the most significant rate gains at high SNR. In this simulation, the increase $R^*(P) - R_M(P)$ is approximately 0.28 bits per channel use for large P .

IV. ASYMPTOTICALLY OPTIMAL CSI

The numerical example considered in the previous section (see Figure 2) assumes that $\tilde{V}(\hat{H})$ and \hat{H} do not depend on the SNR P/N_0 . However, in practical communication systems, the channel estimation error—as measured by the mean error variance $\mathbb{E}[\tilde{V}(\hat{H})]$ —typically decreases as the SNR increases. In this section, we investigate the high-SNR behavior of the rates achievable with and without rate splitting when $\mathbb{E}[\tilde{V}(\hat{H})]$ vanishes as the SNR tends to infinity.

A. Asymptotic Tightness

Let us consider a family of joint distributions of (\hat{H}, \tilde{H}) parametrized by $\rho = P/N_0$. To make this dependence on ρ explicit, we shall write the two channel components as \hat{H}_ρ and \tilde{H}_ρ , and the respective variances as \hat{V}_ρ and $\tilde{V}_\rho(\hat{H}_\rho)$. Similarly, we shall write the entropy power, defined in (36), as $\tilde{\Phi}_\rho(\hat{H}_\rho)$. We further adapt the notation to express Médard's lower bound, the rate-splitting lower bounds (26) and (29), and the upper bounds (34) and (35) as functions of ρ , namely, $R_M(\rho)$, $R^*(\rho, L)$, $R^*(\rho)$, $C_{\text{coh}}(\rho)$, and $I_{\text{upper}}(\rho)$.

We assume that $H = \hat{H}_\rho + \tilde{H}_\rho$ does not depend on ρ and is normalized:

$$\mathbb{E}[|\hat{H}_\rho|^2] + \mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)] = 1. \quad (37)$$

We further assume that the variance of the estimation error \tilde{H}_ρ is not larger than the variance of H , i.e., $\tilde{V}_\rho(\hat{h}_\rho) \leq 1$ for every $\hat{h}_\rho \in \mathbb{C}$.

Theorem 5. *Let \hat{H}_ρ , $\tilde{V}_\rho(\hat{H}_\rho)$ and $\tilde{\Phi}_\rho(\hat{H}_\rho)$ satisfy*

$$\lim_{\rho \rightarrow \infty} \mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)] = 0 \quad (38a)$$

$$\overline{\lim}_{\rho \rightarrow \infty} \mathbb{E}[|\hat{H}_\rho|^4] < \infty \quad (38b)$$

$$\overline{\lim}_{\rho \rightarrow \infty} \left\{ \sup_{\xi \in \mathbb{C}} \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right\} \leq M \quad (38c)$$

for some finite constant M , where we define $0/0 \triangleq 1$ and $a/0 \triangleq \infty$ for every $a > 0$. Then, we have

$$\overline{\lim}_{\rho \rightarrow \infty} \left\{ I(X_G; Y | \hat{H}_\rho) - R^*(\rho) \right\} \leq \log(M). \quad (39)$$

Proof: See Section VII. ■

If \tilde{H}_ρ is Gaussian, then we have $\tilde{V}_\rho(\hat{h}_\rho) = \tilde{\Phi}_\rho(\hat{h}_\rho)$ for $\hat{h}_\rho \in \mathbb{C}$ and the choice $M = 1$ satisfies (38c). Thus, for Gaussian fading the lower bound $R^*(\rho)$ is asymptotically tight:

Corollary 6. *Conditioned on every $\hat{H}_\rho = \hat{h}_\rho$, let \tilde{H}_ρ be Gaussian, and let $\lim_{\rho \rightarrow \infty} \mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)] = 0$. Then, we have*

$$\lim_{\rho \rightarrow \infty} \left\{ I(X_G; Y | \hat{H}_\rho) - R^*(\rho) \right\} = 0. \quad (40)$$

In [5], it was argued that the difference between $R_M(\rho)$ and $C_{\text{coh}}(\rho)$ (34) vanishes as ρ tends to infinity only if $\tilde{V}_\rho(\hat{H}_\rho)$ decays faster than the reciprocal of ρ . In this case, Médard's lower bound is asymptotically tight. In fact, if the fading is Gaussian, then it can be shown that

$$\lim_{\rho \rightarrow \infty} \left\{ I(X_G; Y | \hat{H}_\rho) - R_M(\rho) \right\} = 0 \iff \lim_{\rho \rightarrow \infty} \rho \mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)] = 0. \quad (41)$$

This is in stark contrast to $R^*(\rho)$: Corollary 6 demonstrates that, for Gaussian fading, the lower bound $R^*(\rho)$ is asymptotically tight as long as $\mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)]$ vanishes as ρ tends to infinity, irrespective of the rate of decay.

B. Prediction- and Interpolation-Based Channel Estimation

We evaluate the lower bounds $R_M(\rho)$, $R^*(\rho, 2)$, and $R^*(\rho)$ together with the upper bound $I_{\text{upper}}(\rho)$ for two specific channel estimation errors satisfying (38a). We assume that \hat{H}_ρ and \tilde{H}_ρ are zero-mean, circularly-symmetric Gaussian random variables that are independent of each other, the former with variance \hat{V}_ρ and the latter with variance \tilde{V}_ρ . We consider variances \tilde{V}_ρ of the forms

$$\tilde{V}_\rho = \left(\frac{1}{2B} + \frac{1}{\rho} \right)^{2B} \rho^{2B-1} - \frac{1}{\rho} \quad (42a)$$

and

$$\tilde{V}_\rho = \frac{2BT}{\rho + 2BT} \quad (42b)$$

for some $0 < B < \frac{1}{2}$, where $T = \lfloor 1/(2B) \rfloor$ is the largest integer not greater than $1/(2B)$.

As we shall argue next, (42a) corresponds to prediction-based channel estimation, whereas (42b) corresponds to interpolation-based channel estimation:

Suppose for a moment that the fading process $\{H[k]\}_{k \in \mathbb{Z}}$ is not i.i.d. (as assumed in Section I) but is a zero-mean, unit-variance, stationary, circularly-symmetric, complex Gaussian process with power spectral density

$$f_H(\lambda) = \begin{cases} \frac{1}{2B}, & |\lambda| < B \\ 0, & B \leq |\lambda| \leq \frac{1}{2} \end{cases} \quad (43)$$

for some $0 < B < \frac{1}{2}$. The fading's autocovariance function is determined by $f_H(\cdot)$ through the expression

$$\mathbb{E}[(H[k+m] - \mu)(H[k] - \mu)^*] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda} f_H(\lambda) d\lambda \quad (44)$$

where $(\cdot)^*$ denotes complex conjugation and $i \triangleq \sqrt{-1}$.

We obtain (42a) if we let $\hat{H}[k]$ be the minimum mean-square error (MMSE) predictor in predicting $H[k]$ from a noisy observation of its past

$$H[k-1]\sqrt{P} + Z[k-1], H[k-2]\sqrt{P} + Z[k-2], \dots \quad (45)$$

Indeed, in this case $\hat{H}[k]$ and $\tilde{H}[k] = H[k] - \hat{H}[k]$ are zero-mean, circularly-symmetric Gaussian random variables that are independent of each other, the latter with mean zero and variance [9], [10, Equation (11)]

$$\tilde{V}_\rho = \exp \left\{ \int_{-1/2}^{1/2} \log \left(f_H(\lambda) + \frac{1}{\rho} \right) d\lambda \right\} - \frac{1}{\rho}. \quad (46)$$

For the power spectral density (43) this gives (42a). Note that, even though the lower bounds $R_M(\rho)$, $R^*(\rho, L)$, and $R^*(\rho)$ were derived for i.i.d. fading $\{\hat{H}_\rho[k], \tilde{H}_\rho[k]\}_{k \in \mathbb{Z}}$, by evaluating them for $\tilde{H}_\rho[k]$ having variance (46), they can be used to derive lower bounds on the capacity of noncoherent fading channels with stationary fading having power spectral density $f_H(\cdot)$; see [10].

The variance (42b) corresponds to a channel-estimation scheme where the transmitter emits every T time instants (say at $k = nT$, $n \in \mathbb{Z}$) a pilot symbol \sqrt{P} and where the receiver estimates the fading coefficients at the remaining time instants k (i.e., where k is not an integer multiple of T) from the noisy observations

$$H[nT]\sqrt{P} + Z[nT], \quad n \in \mathbb{Z} \quad (47)$$

using an MMSE interpolator; see, e.g., [11]–[13]. When the power spectral density $f_H(\cdot)$ is bandlimited to B and when $T \leq 1/(2B)$, it can be shown that the variance of the estimation error is given by [14]

$$\tilde{V}_\rho = 1 - \int_{-B}^B \frac{\rho [f_H(\lambda)]^2}{\rho f_H(\lambda) + T} d\lambda. \quad (48)$$

For the power spectral density (43) this gives (42b). Again, even though the lower bounds $R_M(\rho)$, $R^*(\rho, L)$, and $R^*(\rho)$ were derived for i.i.d. fading $\{\hat{H}_\rho[k], \tilde{H}_\rho[k]\}_{k \in \mathbb{Z}}$, by evaluating them for $\tilde{H}_\rho[k]$ having variance (42b), they can be directly used to derive lower bounds on the capacity of noncoherent fading channels with stationary fading having power spectral density $f_H(\cdot)$, provided that we account for the rate loss due to the transmission of pilots. In fact, it was shown that, when $1/(2B)$ is an integer, the above interpolation-based channel estimation scheme together with Médard's lower bound $R_M(\rho)$ achieves the capacity pre-log [12], [13].⁴

C. Numerical Examples

Figure 3(a) shows the lower bounds $R_M(\rho)$, $R^*(\rho, 2)$, and $R^*(\rho)$ together with the upper bounds $I_{\text{upper}}(\rho)$ and $C_{\text{coh}}(\rho)$ as a function of ρ for \tilde{H}_ρ having variance (42a), with $B = 1/4$. Figure 3(b) shows the same bounds as a function of the energy per information bit. The shaded area constitutes the area where the mutual information corresponding to Gaussian inputs may lie. Observe that, in contrast to the curves in Figure 2, all curves are unbounded in the SNR, which is a consequence of the fact that \tilde{V}_ρ vanishes as ρ tends to infinity. Further observe that the shaded area decreases as ρ grows. This is consistent with Corollary 6, which states that for Gaussian fading the lower bound $R^*(\rho)$ is asymptotically tight.

Figure 4(a) shows the lower bounds $R_M(\rho)$, $R^*(\rho, 2)$, and $R^*(\rho)$ together with the upper bounds $I_{\text{upper}}(\rho)$ and $C_{\text{coh}}(\rho)$ as a function of ρ for \tilde{H}_ρ having variance (42b), with $BT = 1/2$. Again, observe that all curves are

⁴The *capacity pre-log* is defined as the limiting ratio of the capacity to $\log(\rho)$ as ρ tends to infinity. In multiple-input multiple-output (MIMO) systems, it is sometimes also referred to as the *number of degrees of freedom* or the *multiplexing gain*.

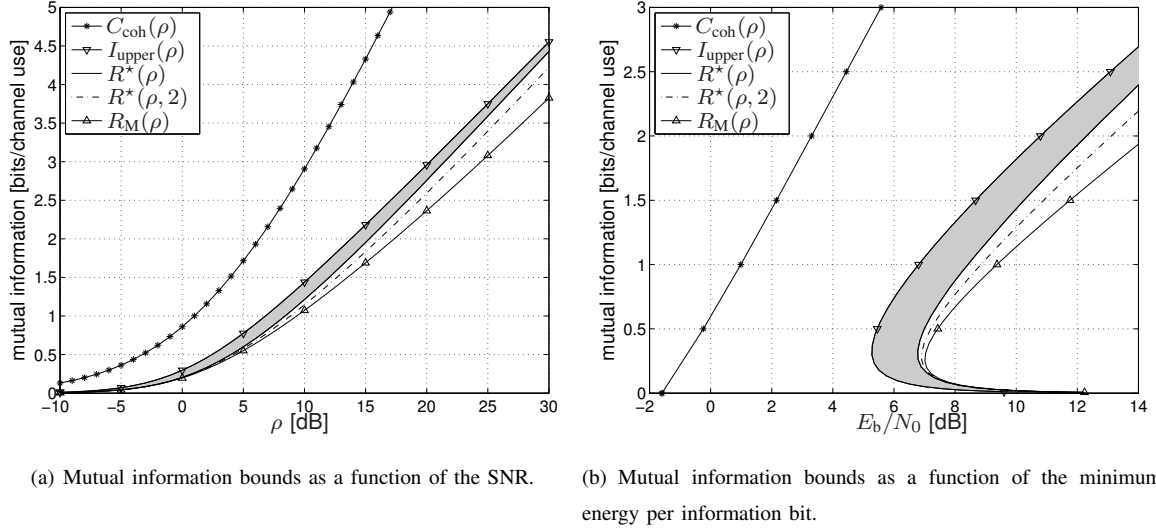


Fig. 3. Prediction-based channel estimation

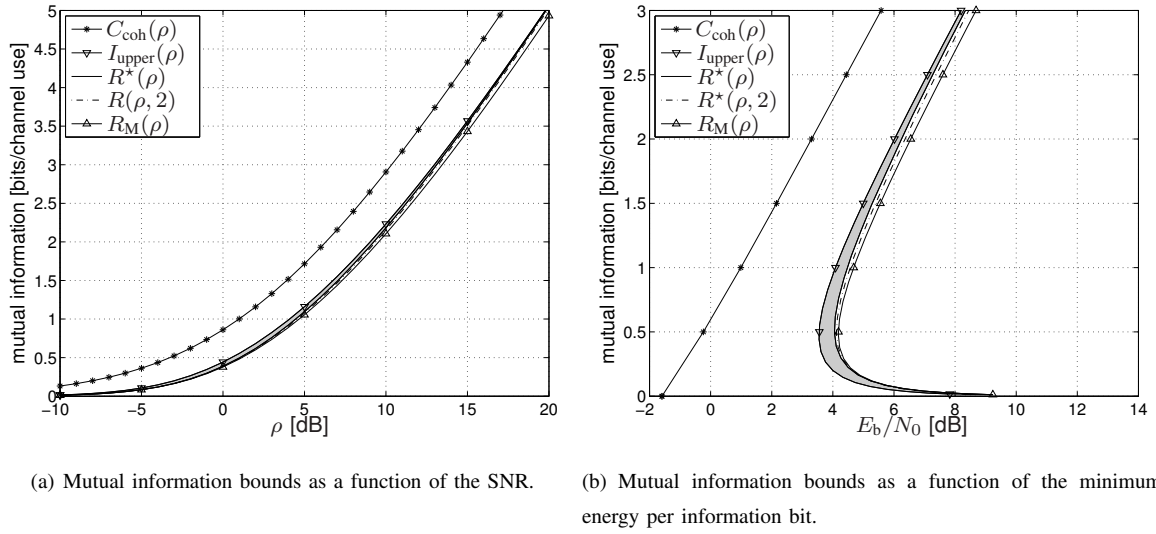


Fig. 4. Interpolation-based channel estimation

unbounded in the SNR and that the lower bound $R^*(\rho)$ is asymptotically tight as ρ tends to infinity. In fact, $R^*(\rho)$ is close to $I_{\text{upper}}(\rho)$ for a large range of SNR. Further observe that, at high SNR, all upper and lower bounds have the same logarithmic slope. This fact was used in [12], [13] to derive tight lower bounds on the capacity pre-log of noncoherent fading channels.

Fig. 5 shows the lower bounds $R_M(\rho)$, $R^*(\rho, 2)$, and $R^*(\rho)$ together with the upper bounds $I_{\text{upper}}(\rho)$ and $C_{\text{coh}}(\rho)$ normalized by $R_M(P)$ as a function of ρ for \tilde{H}_ρ having variance (42b). Observe that, as ρ tends to zero, the ratios of the lower bounds $R^*(\rho, 2)$ and $R^*(\rho)$ to $R_M(P)$ tend to one. Thus, at low SNR, rate splitting only provides moderate rate gains.

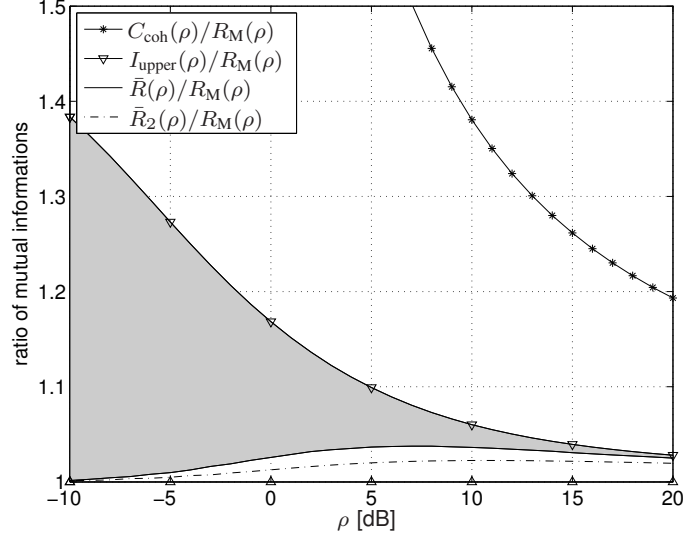


Fig. 5. Mutual information bounds divided by $R_M(\rho)$ as a function of the SNR.

V. RELATIONSHIP TO MISMATCHED DECODING

We have demonstrated that Médard's lower bound $R_M(P)$ on the capacity of fading channels with imperfect CSI can be sharpened by using a rate-splitting approach: by expressing the Gaussian input X as the sum of L Gaussian random variables X_1, \dots, X_L , by applying the chain rule for mutual information to express $I(X; Y | \hat{H})$ as

$$I(X; Y | \hat{H}) = \sum_{\ell=1}^L I(X_\ell; Y | X^{\ell-1}, \hat{H}) \quad (49)$$

and by lower-bounding each mutual information on the RHS of (49) using Médard's bounding technique, we obtain a lower bound that is strictly larger than $R_M(P)$.

This result is reminiscent of a result in the mismatched decoding literature. Indeed, it has been shown that Médard's lower bound $R_M(P)$ is the *generalized mutual information (GMI)*⁵ [15]–[17] of the above channel (1) when the codebook is drawn according to an i.i.d. Gaussian distribution and when the decoding rule is the *scaled nearest neighbor decoding rule* under which the decoder chooses the message m that minimizes [5, Corollary 3.0.1]

$$D(m) = \frac{1}{n} \sum_{k=1}^n \left| y[k] - \hat{h}[k] x^{(m)}[k] \right|^2. \quad (50)$$

Here, $(x^{(m)}[1], \dots, x^{(m)}[n])$ denotes the codeword associated with the message $m \in \{1, \dots, \lfloor e^{nR} \rfloor\}$, and R and n denote the rate and the blocklength of the code, respectively. It has been further shown that, for a given decoding rule, treating the single-user channel as a multiple-access channel (MAC) can sometimes yield an achievable rate

⁵For a given channel and decoding rule, the GMI is the rate below which the average probability of error—averaged over the ensemble of i.i.d. codebooks—decays to zero as the blocklength tends to infinity, and above which this average tends to one.

that is larger than the GMI or other achievable rates corresponding to codebooks under which the codewords are drawn independently [18].

Since the above rate-splitting approach treats the single-user channel (1) as an L -user MAC with channel inputs X_1, \dots, X_L , i.e.,

$$Y = \sum_{\ell=1}^L (\hat{H} + \tilde{H})X_\ell + Z \quad (51)$$

it may therefore seem plausible that said approach can sharpen Médard's lower bound. However, note that, in contrast to [18] where the decoding rule is held fixed and the gain in achievable rate is due to the fact that treating the single-user channel as a MAC allows us to consider codebooks under which the codewords are not necessarily drawn independently, here the sharpening of the lower bound is due to a change of the decoding rule. For example, treating (1) as a two-user MAC (51), the rate $R_2(P_1, P_2)$ corresponding to two-layer rate splitting [cf. (14)] can be achieved by dividing the message m into the submessages (m_1, m_2) , by drawing for each submessage a codebook according to an i.i.d. Gaussian distribution, and by employing a decoder that chooses the pair (m_1, m_2) that minimizes

$$D_1(m_1) = \frac{1}{n} \sum_{k=1}^n \left| y[k] - \hat{h}[k]x_1^{(m_1)}[k] \right|^2 \quad (52)$$

and

$$D_2(m_2|\hat{m}_1) = \frac{1}{n} \sum_{k=1}^n \left| y[k] - \hat{h}[k]x_1^{(\hat{m}_1)}[k] - \hat{h}[k]x_2^{(m_2)}[k] \right|^2. \quad (53)$$

Here $x_\ell^{(m_\ell)}[1], \dots, x_\ell^{(m_\ell)}[n]$, $\ell = 1, 2$ denotes the codeword associated with message m_ℓ , and \hat{m}_1 denotes the submessage that minimizes $D_1(m_1)$. While treating the single-user channel (1) as a MAC (51) gives rise to a codebook under which codewords

$$x_1^{(m_1)}[1] + x_2^{(m_2)}[1], \dots, x_1^{(m_1)}[n] + x_2^{(m_2)}[n]$$

corresponding to different messages $m = (m_1, m_2)$ are not drawn independently, in contrast to [18], this is not the reason why the achievable rate is increased. In fact, it can be shown that the same codebook together with the scaled nearest neighbor decoding rule

$$D(m_1, m_2) = \frac{1}{n} \sum_{k=1}^n \left| y[k] - \hat{h}[k]x_1^{(m_1)}[k] - \hat{h}[k]x_2^{(m_2)}[k] \right|^2 \quad (54)$$

yields an achievable rate that is not larger than $R_M(P)$.

In a nutshell, Médard's lower bound $R_M(P)$ corresponds to the GMI for the scaled nearest neighbor decoding rule (54), whereas the rate-splitting lower bounds $R[\mathbf{Q}]$ correspond to a recursive decoding rule as in (52) and (53). The results from Sections III, and specifically Lemma 2, demonstrate that this recursive decoding rule yields a larger achievable rate than the scaled nearest neighbor decoding rule.

VI. PROOF OF THEOREM 4

To prove Theorem 4, we shall first show that it suffices to consider *uniform layerings*

$$\mathbf{U}(P, K) \triangleq \left(\frac{P}{K}, 2\frac{P}{K}, \dots, (K-1)\frac{P}{K}, P \right). \quad (55)$$

Specifically, we shall show that for every L -layering $\mathbf{Q} \in \mathcal{Q}(P, L)$ there exists some K such that $\mathbf{U}(P, K)$ outperforms \mathbf{Q} , i.e.,

$$R[\mathbf{U}(P, K)] > R[\mathbf{Q}]. \quad (56)$$

This then implies that

$$R^*(P) = \sup_{L \in \mathbb{N}} \left\{ \sup_{\mathbf{Q} \in \mathcal{Q}(P, L)} R[\mathbf{Q}] \right\} = \sup_{K \in \mathbb{N}} R[\mathbf{U}(P, K)] \quad (57)$$

from which we obtain

$$R^*(P) = \overline{\lim}_{K \rightarrow \infty} R[\mathbf{U}(P, K)] \quad (58)$$

upon noting that $\mathbf{U}(P, K) \subset \mathbf{U}(P, 2K)$ for every $k \in \mathbb{N}$.

To show that for every L -layering $\mathbf{Q} \in \mathcal{Q}(P, L)$ there exists some $\mathbf{U}(P, K)$ outperforming \mathbf{Q} , we first note that one can find two $(L+1)$ -layerings $\mathbf{S} \in \mathcal{Q}(P, L+1)$ and $\mathbf{T} \in \mathcal{Q}(P, L+1)$ satisfying $\mathbf{Q} \subset \mathbf{S}$ and $\mathbf{T} \subset \mathbf{U}(P, K)$ such that for every $\epsilon > 0$ and sufficiently large K , we have

$$\max_{1 \leq \ell \leq L+1} |T_\ell - S_\ell| \leq \epsilon. \quad (59)$$

Indeed, \mathbf{S} may be obtained by including $(Q_1 + Q_2)/2$ into \mathbf{Q} , i.e., $\mathbf{S} = \mathbf{Q} \cup \{(Q_1 + Q_2)/2\}$. Furthermore, for K larger than $P/(\min_{1 \leq \ell \leq L} |S_{\ell+1} - S_\ell|)$, choosing

$$T_\ell = \left\lceil \frac{S_\ell K}{P} \right\rceil \frac{P}{K}, \quad \ell = 1, \dots, L+1$$

(where $\lceil x \rceil$ denotes the smallest integer larger than x) yields $\mathbf{T} \subset \mathbf{U}(P, K)$ and

$$\max_{1 \leq \ell \leq L+1} |T_\ell - S_\ell| \leq \frac{P}{K} \quad (60)$$

from which (59) follows. To prove (56), we then need the following lemma.

Lemma 7. *The function $R[\mathbf{Q}]$ satisfies*

$$\lim_{\mathbf{Q} \rightarrow \mathbf{Q}'} R[\mathbf{Q}] = R[\mathbf{Q}'] \quad (61)$$

where $\mathbf{Q} \rightarrow \mathbf{Q}'$ is to be understood as $\max_\ell |Q_\ell - Q'_\ell| \rightarrow 0$.

Proof: See Appendix C. ■

From Lemma 7 and from the observation (59), it follows that for every $\delta > 0$ there exists a sufficiently large K such that

$$|R[\mathbf{T}] - R[\mathbf{S}]| \leq \delta. \quad (62)$$

Since, by Lemma 2, we have

$$R[\mathbf{Q}] < R[\mathbf{S}] \quad \text{and} \quad R[\mathbf{T}] < R[\mathbf{U}(P, K)] \quad (63)$$

this yields

$$R[\mathbf{Q}] < R[\mathbf{S}] < R[\mathbf{T}] + \delta \quad (64)$$

which for a sufficiently small δ is strictly smaller than $R[\mathbf{U}(P, K)]$. This proves (56).

Recalling that (56) implies (58), we continue by evaluating $R[\mathbf{U}(P, K)]$ in the limit as K tends to infinity. To this end, we write $R[\mathbf{U}(P, K)]$ as

$$R[\mathbf{U}(P, K)] = \sum_{\ell=1}^K \mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{U}}(W_{\ell}, \hat{H})) \right] \quad (65)$$

with [cf. (22)]

$$\begin{aligned} \Gamma_{\ell, \mathbf{U}}(W_{\ell}, \hat{H}) &= \frac{|\hat{H}|^2 \frac{P}{K}}{\tilde{V}(\hat{H})(\ell-1) \frac{P}{K} W_{\ell} + \tilde{V}(\hat{H}) \frac{P}{K} + (|\hat{H}|^2 + \tilde{V}(\hat{H}))(P - \ell \frac{P}{K}) + N_0} \\ &= \frac{|\hat{H}|^2}{\tilde{V}(\hat{H})(\ell-1) W_{\ell} + \tilde{V}(\hat{H}) + (|\hat{H}|^2 + \tilde{V}(\hat{H}))(K - \ell) + N_0 \frac{K}{P}} \end{aligned} \quad (66)$$

and

$$W_{\ell} \triangleq \begin{cases} 0, & \ell = 0 \\ \frac{1}{(\ell-1) \frac{P}{K}} |\sum_{i < \ell} X_i|^2, & \ell = 2, \dots, K. \end{cases} \quad (67)$$

The random variables (W_1, \dots, W_K) are dependent but have equal marginals. (Each marginal has a unit-mean exponential distribution.) Since the RHS of (65) depends on (W_1, \dots, W_K) only via the marginal distributions, we can thus express $R[\mathbf{U}(P, K)]$ as

$$R[\mathbf{U}(P, K)] = \mathbb{E} \left[\sum_{\ell=1}^K \log(1 + \Gamma_{\ell, \mathbf{U}}(W, \hat{H})) \right], \quad (68)$$

where W is independent of \hat{H} and has a unit-mean exponential distribution.

Combining (68) with (58) yields

$$R^*(P) = \overline{\lim}_{K \rightarrow \infty} \mathbb{E} \left[\sum_{\ell=1}^K \log(1 + \Gamma_{\ell, \mathbf{U}}(W, \hat{H})) \right]. \quad (69)$$

We next show that

$$R^*(P) = \mathbb{E} \left[\lim_{K \rightarrow \infty} \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(W, \hat{H}) \right] \quad (70)$$

and evaluate $\sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(W, \hat{H})$ for every $(W, \hat{H}) = (w, \hat{h})$ in the limit as K tends to infinity. To this end, we first lower-bound $R^*(P)$ using Fatou's lemma [19] and the lower bound $\log(1+x) \geq x - x^2/2$, $x \geq 0$:

$$\begin{aligned} R^*(P) &= \overline{\lim}_{K \rightarrow \infty} \mathbb{E} \left[\sum_{\ell=1}^K \log(1 + \Gamma_{\ell, \mathbf{U}}(W, \hat{H})) \right] \\ &\geq \mathbb{E} \left[\lim_{K \rightarrow \infty} \sum_{\ell=1}^K \log(1 + \Gamma_{\ell, \mathbf{U}}(W, \hat{H})) \right] \\ &\geq \mathbb{E} \left[\lim_{K \rightarrow \infty} \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(W, \hat{H}) \right] - \frac{1}{2} \mathbb{E} \left[\overline{\lim}_{K \rightarrow \infty} \sum_{\ell=1}^K [\Gamma_{\ell, \mathbf{U}}(W, \hat{H})]^2 \right] \end{aligned} \quad (71)$$

where $\underline{\lim}$ denotes the *limit inferior*. It follows that the second term on the RHS of (71) is zero. Indeed, we have for every $(W, \hat{H}) = (w, \hat{h})$

$$\begin{aligned} \left[\Gamma_{\ell, \mathbf{U}}(w, \hat{h}) \right]^2 &= \frac{|\hat{h}|^4}{\left[\tilde{V}(\hat{h})(\ell - 1)w + \tilde{V}(\hat{h}) + (|\hat{h}|^2 + \tilde{V}(\hat{h}))(K - \ell) + N_0 \frac{K}{P} \right]^2} \\ &\leq \frac{|\hat{h}|^4}{\left[\min \left\{ \tilde{V}(\hat{h})w, (|\hat{h}|^2 + \tilde{V}(\hat{h})) \right\} (K - 1) + \tilde{V}(\hat{h}) + N_0 \frac{K}{P} \right]^2}, \end{aligned} \quad (72)$$

from which we obtain

$$\sum_{\ell=1}^K \left[\Gamma_{\ell, \mathbf{U}}(w, \hat{h}) \right]^2 \leq \frac{K |\hat{h}|^4}{\left[\min \left\{ \tilde{V}(\hat{h})w, (|\hat{h}|^2 + \tilde{V}(\hat{h})) \right\} (K - 1) + \tilde{V}(\hat{h}) + N_0 \frac{K}{P} \right]^2}. \quad (73)$$

Since $\sum_{\ell=1}^K \left[\Gamma_{\ell, \mathbf{U}}(w, \hat{h}) \right]^2$ is nonnegative, and since the RHS of (73) vanishes as K tends to infinity, it follows that, for every $(W, \hat{H}) = (w, \hat{h})$,

$$\overline{\lim}_{K \rightarrow \infty} \sum_{\ell=1}^K \left[\Gamma_{\ell, \mathbf{U}}(w, \hat{h}) \right]^2 = 0. \quad (74)$$

Combining (74) with (71) yields

$$R^*(P) \geq \mathbb{E} \left[\overline{\lim}_{K \rightarrow \infty} \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(W, \hat{H}) \right]. \quad (75)$$

We next show that

$$R^*(P) \leq \mathbb{E} \left[\overline{\lim}_{K \rightarrow \infty} \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(W, \hat{H}) \right]. \quad (76)$$

To this end, we first use the upper bound $\log(1 + x) \leq x$, $x \geq 0$ to obtain

$$\begin{aligned} R^*(P) &= \overline{\lim}_{K \rightarrow \infty} \mathbb{E} \left[\sum_{\ell=1}^K \log(1 + \Gamma_{\ell, \mathbf{U}}(W, \hat{H})) \right] \\ &\leq \overline{\lim}_{K \rightarrow \infty} \mathbb{E} \left[\sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(W, \hat{H}) \right]. \end{aligned} \quad (77)$$

Noting that, for every $(W, \hat{H}) = (w, \hat{h})$, the sum inside the expectation is upper-bounded by

$$\sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(w, \hat{h}) \leq \sum_{\ell=1}^K \frac{|\hat{h}|^2}{N_0 \frac{K}{P}} = \frac{P |\hat{h}|^2}{N_0} \triangleq \zeta(\hat{h}) \quad (78)$$

and noting that, since \hat{H} has a finite second moment, we have that $0 < \mathbb{E}[\zeta(\hat{H})] < \infty$, we obtain (76) upon applying Fatou's lemma to the function $(w, \hat{h}) \mapsto \zeta(\hat{h}) - \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(w, \hat{h})$.

It remains to show that, for every $(W, \hat{H}) = (w, \hat{h})$,

$$\lim_{K \rightarrow \infty} \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(w, \hat{h}) = \frac{|\hat{h}|^2}{|\hat{h}|^2 + \tilde{V}(\hat{h}) + \frac{N_0}{P}} \Theta \left(\frac{\tilde{V}(\hat{h})(w - 1) - |\hat{h}|^2}{|\hat{h}|^2 + \tilde{V}(\hat{h}) + \frac{N_0}{P}} \right) \quad (79)$$

where $\Theta(\cdot)$ is defined in (31). This then implies that (75) and (76) coincide and

$$R^*(P) = \mathbb{E} \left[\frac{|\hat{H}|^2}{|\hat{H}|^2 + \tilde{V}(\hat{H}) + \frac{N_0}{P}} \Theta \left(\frac{\tilde{V}(\hat{H})(W - 1) - |\hat{H}|^2}{|\hat{H}|^2 + \tilde{V}(\hat{H}) + \frac{N_0}{P}} \right) \right] \quad (80)$$

which proves Theorem 4.

To show (79), we express the denominator in $\Gamma_{\ell, \mathbf{U}}(w, \hat{h})$ as $a\ell + bK + c$ with

$$a = \tilde{V}(\hat{h})(w - 1) - |\hat{h}|^2 \quad (81a)$$

$$b = |\hat{h}|^2 + \tilde{V}(\hat{h}) + \frac{N_0}{P} \quad (81b)$$

$$c = \tilde{V}(\hat{h})(1 - w) \quad (81c)$$

allowing us to write

$$\sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(w, \hat{h}) = \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK + c}. \quad (82)$$

If $a = 0$, then this becomes

$$\lim_{K \rightarrow \infty} \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(w, \hat{h}) = \frac{|\hat{h}|^2}{b}. \quad (83)$$

We next consider the case $a \neq 0$. Note that

$$\lim_{K \rightarrow \infty} \left(\sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK + c} - \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK} \right) = 0. \quad (84)$$

Indeed, by the triangle inequality, we have

$$\left| \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK + c} - \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK} \right| \leq \sum_{\ell=1}^K \frac{|c| |\hat{h}|^2}{(a\ell + bK + c)(a\ell + bK)}. \quad (85)$$

Since the two factors $(a\ell + bK + c)$ and $(a\ell + bK)$ appearing in the denominator are both positive affine functions of k with equal coefficient a , their product takes its extremal values at $k = 1$ or $k = K$, depending on the sign of a . If $a > 0$, then

$$\left| \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK + c} - \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK} \right| \leq \frac{K|c| |\hat{h}|^2}{(a + bK + c)(a + bK)}. \quad (86)$$

If $a \leq 0$, then

$$\left| \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK + c} - \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK} \right| \leq \frac{K|c| |\hat{h}|^2}{((a + b)K + c)(a + b)K}. \quad (87)$$

Since the RHS of (86) and of (87) vanish as K tends to infinity, this yields (84). Consequently,

$$\begin{aligned} \lim_{K \rightarrow \infty} \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK + c} &= \lim_{K \rightarrow \infty} \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a\ell + bK} \\ &= \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{\ell=1}^K \frac{|\hat{h}|^2}{a \frac{\ell}{K} + b} \\ &= \int_0^1 \frac{|\hat{h}|^2}{ax + b} dx \\ &= \frac{|\hat{h}|^2}{a} \log \left(1 + \frac{a}{b} \right) \end{aligned} \quad (88)$$

where the third step follows by noting that the function $x \mapsto \frac{1}{ax+b}$ is Riemann integrable, so the Riemann sum converges to the integral.

Using the definition of $\Theta(\cdot)$, it follows from (83) and (88) that

$$\lim_{K \rightarrow \infty} \sum_{\ell=1}^K \Gamma_{\ell, \mathbf{U}}(w, \hat{h}) = \frac{|\hat{h}|^2}{b} \Theta\left(\frac{a}{b}\right) = \frac{|\hat{h}|^2}{|\hat{h}|^2 + \tilde{V}(\hat{h}) + \frac{N_0}{P}} \Theta\left(\frac{\tilde{V}(\hat{h})(w-1) - |\hat{h}|^2}{|\hat{h}|^2 + \tilde{V}(\hat{h}) + \frac{N_0}{P}}\right) \quad (89)$$

thus proving (79), which in turn proves Theorem 4.

VII. PROOF OF THEOREM 5

To prove Theorem 5, we show that, in the limit as the SNR tends to infinity, the difference

$$I(X_G; Y | \hat{H}_\rho) - R^*(\rho) \quad (90)$$

is upper-bounded by $\log(M)$ provided that (38b)–(38c) are satisfied. We express $R^*(\rho)$ as $\mathbb{E}[R^*(\rho, W, \hat{H})]$ with

$$R^*(\rho, w, \xi) \triangleq \frac{|\xi|^2}{|\xi|^2 + \tilde{V}_\rho(\xi) + \rho^{-1}} \Theta\left(\frac{(w-1)\tilde{V}_\rho(\xi) - |\xi|^2}{|\xi|^2 + \tilde{V}_\rho(\xi) + \rho^{-1}}\right), \quad (\rho > 0, w \geq 0, \xi \in \mathbb{C}) \quad (91)$$

and upper-bound $I(X_G; Y | \hat{H}_\rho)$ using (35):

$$\begin{aligned} I(X_G; Y | \hat{H}_\rho) - R^*(\rho) &\leq \mathbb{E}[R_M(\rho, \hat{H}_\rho)] + \mathbb{E}[\Delta(\rho, W, \hat{H}_\rho)] - \mathbb{E}[R^*(\rho, W, \hat{H}_\rho)] \\ &= \mathbb{E}[\Sigma(\rho, \hat{H}_\rho)] \end{aligned} \quad (92)$$

where we have defined

$$R_M(\rho, \xi) \triangleq \log\left(1 + \frac{|\xi|^2}{\tilde{V}_\rho(\xi) + \rho^{-1}}\right), \quad (\rho > 0, w \geq 0, \xi \in \mathbb{C}) \quad (93)$$

$$\Delta(\rho, w, \xi) \triangleq \log\left(\frac{\tilde{V}_\rho(\xi) + \rho^{-1}}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}}\right), \quad (\rho > 0, w \geq 0, \xi \in \mathbb{C}) \quad (94)$$

and

$$\Sigma(\rho, \xi) \triangleq R_M(\rho, \xi) + \mathbb{E}[\Delta(\rho, W, \xi)] - \mathbb{E}[R^*(\rho, W, \xi)], \quad (\rho > 0, \xi \in \mathbb{C}). \quad (95)$$

We next show that

$$\overline{\lim}_{\rho \rightarrow \infty} \mathbb{E}[\Sigma(\rho, \hat{H}_\rho)] \leq \log(M). \quad (96)$$

To this end, we write (92) as

$$\mathbb{E}[\Sigma(\rho, \hat{H}_\rho)] = \mathbb{E}[\Sigma(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| \leq \xi_0\}] + \mathbb{E}[\Sigma(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| > \xi_0\}] \quad (97)$$

where $\mathbf{I}\{\cdot\}$ denotes the indicator function (it is 1 if the statement in the curly brackets is true and is 0 otherwise).

We then show that

$$\lim_{\xi_0 \downarrow 0} \overline{\lim}_{\rho \rightarrow \infty} \mathbb{E}[\Sigma(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| \leq \xi_0\}] = 0 \quad (98a)$$

and

$$\overline{\lim}_{\xi_0 \downarrow 0} \overline{\lim}_{\rho \rightarrow \infty} \mathbb{E}[\Sigma(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| > \xi_0\}] \leq \log(M). \quad (98b)$$

To prove (98a), we need the following two lemmas.

Lemma 8. As ρ tends to infinity, we have

$$\overline{\lim}_{\rho \rightarrow \infty} \sup_{\xi \in \mathbb{C}} \Sigma(\rho, \xi) \leq \gamma + \log(M). \quad (99)$$

Proof: See Appendix D. ■

Lemma 9. Let $\tilde{V}_\rho(\hat{H}_\rho)$ and \tilde{H}_ρ satisfy (38a) and (38c). Then

$$\lim_{\xi_0 \downarrow 0} \overline{\lim}_{\rho \rightarrow \infty} \Pr\{|\hat{H}_\rho| > \xi_0\} = 1. \quad (100)$$

Proof: See Appendix E. ■

Lemma 8 implies that, for every $\epsilon > 0$ there exists a $\rho_0 > 0$ such that

$$\sup_{\xi \in \mathbb{C}} \Sigma(\rho, \xi) \leq \gamma + \log(M + \epsilon), \quad \rho \geq \rho_0. \quad (101)$$

Consequently, for $\rho \geq \rho_0$ we have

$$\mathbb{E}\left[\Sigma(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| \leq \xi_0\}\right] \leq (\gamma + \log(M + \epsilon)) \Pr\{|\hat{H}_\rho| \leq \xi_0\}. \quad (102)$$

Together with Lemma 9, this yields (98a) upon taking limits on both sides of (102):

$$\begin{aligned} & \lim_{\xi_0 \downarrow 0} \overline{\lim}_{\rho \rightarrow \infty} \left\{ \mathbb{E}\left[\Sigma(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| \leq \xi_0\}\right] \right\} \\ & \leq (\gamma + \log(M + \epsilon)) \left\{ \lim_{\xi_0 \downarrow 0} \overline{\lim}_{\rho \rightarrow \infty} \Pr\{|\hat{H}_\rho| \leq \xi_0\} \right\} \\ & = 0. \end{aligned} \quad (103)$$

To prove (98b), we first upper-bound $\Sigma(\rho, \xi)$ by lower-bounding $\mathbb{E}[R^*(\rho, W, \xi)]$ for $\rho > 0$ and $|\xi| > \xi_0$ using that $R^*(\rho, w, \xi)$ is nonnegative:

$$\mathbb{E}[R^*(\rho, W, \xi)] \geq \int_0^{\kappa(\rho, \xi)} R^*(\rho, w, \xi) e^{-w} dw, \quad (\rho > 0, |\xi| > \xi_0) \quad (104)$$

where

$$\kappa(\rho, \xi) \triangleq \frac{\xi_0^2}{\sqrt{\tilde{V}_\rho(\xi) + \rho^{-1}}}. \quad (105)$$

This choice for $\kappa(\rho, \xi)$ together with the assumption $\tilde{V}_\rho(\xi) \leq 1$ ensures that $(w - 1)\tilde{V}_\rho(\xi) - |\xi|^2$ is negative for all values of the integration variable w and for all $|\xi| > \xi_0$. Using the definition (31) of the function $\Theta(\cdot)$ in (91), the lower bound (104) reads as

$$\mathbb{E}[R^*(\rho, W, \xi)] \geq \int_0^{\kappa(\rho, \xi)} \frac{|\xi|^2}{(w - 1)\tilde{V}_\rho(\xi) - |\xi|^2} \log \left(1 + \frac{(w - 1)\tilde{V}_\rho(\xi) - |\xi|^2}{|\xi|^2 + \tilde{V}_\rho(\xi) + \rho^{-1}} \right) e^{-w} dw, \quad (\rho > 0, |\xi| > \xi_0). \quad (106)$$

Combining (106) with (93)–(95) yields

$$\begin{aligned} \Sigma(\rho, \xi) & \leq \log \left(1 + \frac{|\xi|^2}{\tilde{V}_\rho(\xi) + \rho^{-1}} \right) + \int_0^\infty \log \left(\frac{\tilde{V}_\rho(\xi) + \rho^{-1}}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}} \right) e^{-w} dw \\ & \quad - \int_0^{\kappa(\rho, \xi)} \frac{|\xi|^2}{(w - 1)\tilde{V}_\rho(\xi) - |\xi|^2} \log \left(1 + \frac{(w - 1)\tilde{V}_\rho(\xi) - |\xi|^2}{|\xi|^2 + \tilde{V}_\rho(\xi) + \rho^{-1}} \right) e^{-w} dw, \quad (\rho > 0, |\xi| > \xi_0). \end{aligned} \quad (107)$$

Upon reordering terms, this inequality can be written as

$$\Sigma(\rho, \xi) \leq J_1(\rho, \xi) + J_2(\rho, \xi) + J_3(\rho, \xi) + J_4(\rho, \xi) \quad (108)$$

with

$$J_1(\rho, \xi) \triangleq \int_0^{\kappa(\rho, \xi)} \log \left(\frac{\tilde{V}_\rho(\xi)w + \rho^{-1}}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}} \right) e^{-w} dw \quad (109a)$$

$$J_2(\rho, \xi) \triangleq \int_{\kappa(\rho, \xi)}^\infty \log \left(\frac{\tilde{V}_\rho(\xi) + \rho^{-1}}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}} \right) e^{-w} dw \quad (109b)$$

$$J_3(\rho, \xi) \triangleq \log(|\xi|^2 + \tilde{V}_\rho(\xi) + \rho^{-1}) \int_0^{\kappa(\rho, \xi)} \frac{(1-w)\tilde{V}_\rho(\xi)}{(1-w)\tilde{V}_\rho(\xi) + |\xi|^2} e^{-w} dw \quad (109c)$$

$$J_4(\rho, \xi) \triangleq - \int_0^{\kappa(\rho, \xi)} \frac{(1-w)\tilde{V}_\rho(\xi)}{(1-w)\tilde{V}_\rho(\xi) + |\xi|^2} \log(\tilde{V}_\rho(\xi)w + \rho^{-1}) e^{-w} dw. \quad (109d)$$

We proceed by showing that, for every $\xi_0 > 0$,

$$\lim_{\rho \rightarrow \infty} \mathbb{E}[J_1(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| > \xi_0\}] \leq \log(M) \quad (110a)$$

$$\lim_{\rho \rightarrow \infty} \mathbb{E}[J_i(\rho, \hat{H}_\rho) \mathbf{I}\{|\hat{H}_\rho| > \xi_0\}] \leq 0, \quad i = 2, 3, 4. \quad (110b)$$

The claim (98b) follows then by combining (110a) and (110b) with (108) and by letting ξ_0 tend to zero from above.

To prove (110a) and (110b), the following lemma will be useful.

Lemma 10. *Consider the family of random variables Υ_ρ parametrized by $\rho > 0$ taking values on $(0, 1]$ and satisfying $\lim_{\rho \rightarrow \infty} \mathbb{E}[\Upsilon_\rho] = 0$. Let $f(\cdot)$ be a continuous bounded function on the interval $(0, 1]$ with limit $\lim_{t \downarrow 0} f(t) = f_0$. Then, we have*

$$\lim_{\rho \rightarrow \infty} \mathbb{E}[f(\Upsilon_\rho)] = f_0. \quad (111)$$

Proof: See Appendix F. ■

A. Limit related to $J_1(\rho, \xi)$

Noting that $\tilde{\Phi}_\rho(\xi) \leq \tilde{V}_\rho(\xi)$, we have that

$$w \mapsto \frac{\tilde{V}_\rho(\xi)w + \rho^{-1}}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}}$$

is monotonically increasing in w . Consequently, $J_1(\rho, \xi)$ is upper-bounded by

$$\begin{aligned} J_1(\rho, \xi) &\leq \int_0^{\kappa(\rho, \xi)} \log \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} e^{-w} dw \\ &\leq \left[1 - \exp \left(- \frac{\xi_0^2}{\sqrt{\tilde{V}_\rho(\xi) + \rho^{-1}}} \right) \right] \log \left(\sup_{\xi \in \mathbb{C}} \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right), \quad (\rho > 0, |\xi| > \xi_0). \end{aligned} \quad (112)$$

Averaging (112) over \hat{H}_ρ , and upper-bounding

$$\mathbf{I}\{|\hat{H}_\rho| > \xi_0\} \leq 1 \quad (113)$$

yields

$$\mathbb{E}[J_1(\rho, \hat{H}_\rho) \mathbb{I}\{|\hat{H}_\rho| > \xi_0\}] \leq \mathbb{E} \left[1 - \exp \left(-\frac{\xi_0^2}{\sqrt{\tilde{V}_\rho(\hat{H}_\rho) + \rho^{-1}}} \right) \right] \log \left(\sup_{\xi \in \mathbb{C}} \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right), \quad \rho > 0. \quad (114)$$

Noting that the function $t \mapsto \exp(-\xi_0^2/\sqrt{t})$ is continuous and bounded on $(0, \infty)$ and vanishes as t tends to zero, it follows from (38a) and Lemma 10 that

$$\lim_{\rho \rightarrow \infty} \mathbb{E} \left[\exp \left(-\frac{\xi_0^2}{\sqrt{\tilde{V}_\rho(\hat{H}_\rho) + \rho^{-1}}} \right) \right] = 0. \quad (115)$$

We further have, by (38c) and the continuity of $x \mapsto \log(x)$, that

$$\overline{\lim}_{\rho \rightarrow \infty} \log \left(\sup_{\xi \in \mathbb{C}} \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right) \leq \log(M). \quad (116)$$

Combining (115) and (116) with (114) proves (110a).

B. Limit related to $J_2(\rho, \xi)$

To upper-bound $J_2(\rho, \xi)$, we use that, for $w \geq \kappa(\rho, \xi)$,

$$\begin{aligned} \frac{\tilde{V}_\rho(\xi) + \rho^{-1}}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}} &= \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}} + \frac{\rho^{-1}}{\tilde{\Phi}_\rho(\xi)w + \rho^{-1}} \\ &\leq \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)\kappa(\rho, \xi)} + 1 \\ &\leq \sup_{\xi \in \mathbb{C}} \left\{ \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right\} \frac{\sqrt{1 + \rho^{-1}}}{\xi_0^2} + 1 \end{aligned} \quad (117)$$

where the first inequality follows by lower-bounding $\rho^{-1} \geq 0$ and $w \geq \kappa(\rho, \xi)$ in the denominator of the first fraction and by lower-bounding $\tilde{\Phi}_\rho(\xi)w \geq 0$ in the denominator of the second fraction; and where the second inequality follows by lower-bounding $\kappa(\rho, \xi) \geq \xi_0^2/\sqrt{1 + \rho^{-1}}$ using $\tilde{V}_\rho(\xi) \leq 1$ and by maximizing over ξ . Combining (117) with (109b) yields

$$\begin{aligned} J_2(\rho, \xi) &\leq \log \left(1 + \sup_{\xi \in \mathbb{C}} \left\{ \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right\} \frac{\sqrt{1 + \rho^{-1}}}{\xi_0^2} \right) \int_{\kappa(\rho, \xi)}^{\infty} e^{-w} dw \\ &= \exp \left(-\frac{\xi_0^2}{\sqrt{\tilde{V}_\rho(\hat{H}_\rho) + \rho^{-1}}} \right) \log \left(1 + \sup_{\xi \in \mathbb{C}} \left\{ \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right\} \frac{\sqrt{1 + \rho^{-1}}}{\xi_0^2} \right). \end{aligned} \quad (118)$$

Averaging (118) over \hat{H}_ρ , using (113), and upper-bounding $\log(1 + x) \leq x$ gives

$$\begin{aligned} &\mathbb{E}[J_2(\rho, \hat{H}_\rho) \mathbb{I}\{|\hat{H}_\rho| > \xi_0\}] \\ &\leq \mathbb{E} \left[\exp \left(-\frac{\xi_0^2}{\sqrt{\tilde{V}_\rho(\hat{H}_\rho) + \rho^{-1}}} \right) \right] \sup_{\xi \in \mathbb{C}} \left\{ \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)} \right\} \frac{\sqrt{1 + \rho^{-1}}}{\xi_0^2}, \quad \rho > 0. \end{aligned} \quad (119)$$

Combining (119) with (115) and (38c) proves (110b) for $i = 2$.

C. Limit related to $J_3(\rho, \xi)$

To prove (110b) for $i = 3$, we shall prove the stronger statement

$$\lim_{\rho \rightarrow \infty} \mathbb{E} \left[|J_3(\rho, \hat{H}_\rho)| \mathbb{I}\{|\hat{H}_\rho| > \xi_0\} \right] = 0. \quad (120)$$

To this end, note that, by the triangle inequality,

$$\left| \frac{(1-w)\tilde{V}_\rho(\xi)}{(1-w)\tilde{V}_\rho(\xi) + |\xi|^2} \right| \leq \frac{(1+\kappa(\rho, \xi))\tilde{V}_\rho(\xi)}{(1-\kappa(\rho, \xi))\tilde{V}_\rho(\xi) + \xi_0^2}, \quad 0 \leq w \leq \kappa(\rho, \xi). \quad (121)$$

In (121) we have also used that the denominator $(1-w)\tilde{V}_\rho(\xi) + |\xi|^2$ is positive for $0 \leq w \leq \kappa(\rho, \xi)$ and that $|\xi| \geq \xi_0$, so the denominator is lower-bounded by $(1-\kappa(\rho, \xi))\tilde{V}_\rho(\xi) + \xi_0^2$.

It follows from (121) that the absolute value of the integral in (109c) is upper-bounded by

$$\left| \int_0^{\kappa(\rho, \xi)} \frac{(1-w)\tilde{V}_\rho(\xi)}{(1-w)\tilde{V}_\rho(\xi) + |\xi|^2} e^{-w} dw \right| \leq \left(1 - e^{-\kappa(\rho, \xi)}\right) \frac{(1+\kappa(\rho, \xi))\tilde{V}_\rho(\xi)}{(1-\kappa(\rho, \xi))\tilde{V}_\rho(\xi) + \xi_0^2}. \quad (122)$$

Consequently, we have

$$\begin{aligned} |J_3(\rho, \xi)| &\leq \left(1 - e^{-\kappa(\rho, \xi)}\right) \frac{(1+\kappa(\rho, \xi))\tilde{V}_\rho(\xi)}{(1-\kappa(\rho, \xi))\tilde{V}_\rho(\xi) + \xi_0^2} \left| \log(|\xi|^2 + \tilde{V}_\rho(\xi) + \rho^{-1}) \right| \\ &\leq \frac{(1+\kappa(\rho, \xi))(\tilde{V}_\rho(\xi) + \rho^{-1})}{\xi_0^2 - (\kappa(\rho, \xi) - 1)^+(\tilde{V}_\rho(\xi) + \rho^{-1})} \left| \log(|\xi|^2 + \tilde{V}_\rho(\xi) + \rho^{-1}) \right|, \quad (\rho > 0, |\xi| \geq \xi_0) \end{aligned} \quad (123)$$

where we define $(a)^+ = \max(a, 0)$.⁶ Here the last step follows by upper-bounding $\tilde{V}_\rho(\xi) \leq \tilde{V}_\rho(\xi) + \rho^{-1}$ and by lower-bounding $(1-\kappa(\rho, \xi))\tilde{V}_\rho(\xi) \geq -(\kappa(\rho, \xi) - 1)^+(\tilde{V}_\rho(\xi) + \rho^{-1})$ and $e^{-\kappa(\rho, \xi)} \geq 0$.

Using the definition (105) of $\kappa(\rho, \xi)$, and defining $\mathcal{V}_\rho(\xi) \triangleq \tilde{V}_\rho(\xi) + \rho^{-1}$, the RHS of (123) reads as

$$\frac{\mathcal{V}_\rho(\xi) + \xi_0^2 \sqrt{\mathcal{V}_\rho(\xi)}}{\xi_0^2 - (\xi_0^2 / \sqrt{\mathcal{V}_\rho(\xi)} - 1)^+ \mathcal{V}_\rho(\xi)} \left| \log(|\xi|^2 + \mathcal{V}_\rho(\xi)) \right|. \quad (124)$$

Furthermore, using that $\tilde{V}_\rho(\xi) \leq 1$ and that $x \mapsto \log(x)$ is a monotonically increasing function satisfying $\log(1+x) \leq x$, gives

$$\log(\mathcal{V}_\rho(\xi)) \leq \log(|\xi|^2 + \mathcal{V}_\rho(\xi)) \leq \rho^{-1} + |\xi|^2. \quad (125)$$

The absolute value of the logarithm on the RHS of (124) is thus upper-bounded by

$$\left| \log(|\xi|^2 + \mathcal{V}_\rho(\xi)) \right| \leq \max \left\{ \left| \log(\mathcal{V}_\rho(\xi)) \right|, (\rho^{-1} + |\xi|^2) \right\} \leq \left| \log(\mathcal{V}_\rho(\xi)) \right| + \rho^{-1} + |\xi|^2. \quad (126)$$

By noting that

$$\mathcal{V}_\rho(\xi) \mapsto \frac{\mathcal{V}_\rho(\xi) + \xi_0^2 \sqrt{\mathcal{V}_\rho(\xi)}}{\xi_0^2 - (\xi_0^2 / \sqrt{\mathcal{V}_\rho(\xi)} - 1)^+ \mathcal{V}_\rho(\xi)} \left| \log(\mathcal{V}_\rho(\xi)) \right|$$

⁶The condition $\xi_0 < 1$ ensures that the denominator is still positive.

is a continuous and bounded function of $0 < \Upsilon_\rho(\xi) \leq 1 + \rho^{-1}$ that vanishes as $\Upsilon_\rho(\xi)$ tends to zero, we obtain from (38a) and Lemma 10 that

$$\begin{aligned} & \lim_{\rho \rightarrow \infty} \mathbb{E} \left[\frac{\Upsilon_\rho(\hat{H}_\rho) + \xi_0^2 \sqrt{\Upsilon_\rho(\hat{H}_\rho)}}{\xi_0^2 - \left(\frac{\xi_0^2}{\sqrt{\Upsilon_\rho(\hat{H}_\rho)}} - 1 \right)^+ \Upsilon_\rho(\hat{H}_\rho)} |\log(\Upsilon_\rho(\hat{H}_\rho))| \mathbb{I}\{|\hat{H}_\rho| > \xi_0\} \right] \\ & \leq \lim_{\rho \rightarrow \infty} \mathbb{E} \left[\frac{\Upsilon_\rho(\hat{H}_\rho) + \xi_0^2 \sqrt{\Upsilon_\rho(\hat{H}_\rho)}}{\xi_0^2 - \left(\frac{\xi_0^2}{\sqrt{\Upsilon_\rho(\hat{H}_\rho)}} - 1 \right)^+ \Upsilon_\rho(\hat{H}_\rho)} |\log(\Upsilon_\rho(\hat{H}_\rho))| \right] \\ & = 0 \end{aligned} \tag{127}$$

where the inequality follows from (113). Furthermore, (113) together with the Cauchy-Schwarz inequality yields

$$\begin{aligned} & \mathbb{E} \left[\frac{\Upsilon_\rho(\hat{H}_\rho) + \xi_0^2 \sqrt{\Upsilon_\rho(\hat{H}_\rho)}}{\xi_0^2 - \left(\frac{\xi_0^2}{\sqrt{\Upsilon_\rho(\hat{H}_\rho)}} - 1 \right)^+ \Upsilon_\rho(\hat{H}_\rho)} \left(\rho^{-1} + |\hat{H}_\rho|^2 \right) \mathbb{I}\{|\hat{H}_\rho| > \xi_0\} \right] \\ & \leq \mathbb{E} \left[\frac{\Upsilon_\rho(\hat{H}_\rho) + \xi_0^2 \sqrt{\Upsilon_\rho(\hat{H}_\rho)}}{\xi_0^2 - \left(\frac{\xi_0^2}{\sqrt{\Upsilon_\rho(\hat{H}_\rho)}} - 1 \right)^+ \Upsilon_\rho(\hat{H}_\rho)} \left(\rho^{-1} + |\hat{H}_\rho|^2 \right) \right] \\ & \leq \sqrt{\mathbb{E} \left[\left(\frac{\Upsilon_\rho(\hat{H}_\rho) + \xi_0^2 \sqrt{\Upsilon_\rho(\hat{H}_\rho)}}{\xi_0^2 - \left(\frac{\xi_0^2}{\sqrt{\Upsilon_\rho(\hat{H}_\rho)}} - 1 \right)^+ \Upsilon_\rho(\hat{H}_\rho)} \right)^2 \right]} \sqrt{\mathbb{E} \left[\left(\rho^{-1} + |\hat{H}_\rho|^2 \right)^2 \right]}. \end{aligned} \tag{128}$$

Note that the term inside the first expected value is a continuous and bounded function of $0 < \Upsilon_\rho(\xi) \leq 1 + \rho^{-1}$ that vanishes as $\Upsilon_\rho(\xi)$ tends to zero, so it follows from (38a) and Lemma 10 that the first expected value on the RHS of (128) vanishes as ρ tends to infinity. Further note that, for sufficiently large ρ , the second expectation on the RHS of (128) is bounded since, by (38b),

$$\overline{\lim}_{\rho \rightarrow \infty} \mathbb{E} \left[\left(\rho^{-1} + |\hat{H}_\rho|^2 \right)^2 \right] = \overline{\lim}_{\rho \rightarrow \infty} \mathbb{E} \left[|\hat{H}_\rho|^4 \right] < \infty. \tag{129}$$

Consequently, the above arguments combine to demonstrate that

$$\lim_{\rho \rightarrow \infty} \mathbb{E} \left[\frac{\Upsilon_\rho(\hat{H}_\rho) + \xi_0^2 \sqrt{\Upsilon_\rho(\hat{H}_\rho)}}{\xi_0^2 - \left(\frac{\xi_0^2}{\sqrt{\Upsilon_\rho(\hat{H}_\rho)}} - 1 \right)^+ \Upsilon_\rho(\hat{H}_\rho)} \left(\rho^{-1} + |\hat{H}_\rho|^2 \right) \mathbb{I}\{|\hat{H}_\rho| > \xi_0\} \right] = 0. \tag{130}$$

Combining (130) and (127) with (126) and (123) proves (120).

D. Limit related to $J_4(\rho, \xi)$

To prove (110b) for $i = 4$, first note that $\xi_0 < 1$ implies that, for sufficiently large ρ , we have

$$\tilde{V}_\rho(\xi)w + \rho^{-1} \leq 1, \quad 0 \leq w \leq \kappa(\rho, \xi). \tag{131}$$

Further note that $t \mapsto t/(t + |\xi|^2)$ is monotonically increasing on $(-|\xi|^2, \infty)$. Consequently, for sufficiently large ρ , (109d) is upper-bounded by

$$\begin{aligned} J_4(\rho, \xi) &\leq -\frac{\tilde{V}_\rho(\xi)}{\tilde{V}_\rho(\xi) + |\xi|^2} \int_0^{\kappa(\rho, \xi)} \log(\tilde{V}_\rho(\xi)w + \rho^{-1}) e^{-w} dw \\ &\leq \frac{\tilde{V}_\rho(\xi)}{\tilde{V}_\rho(\xi) + |\xi|^2} \left[\left(1 - e^{-\kappa(\xi, \rho)}\right) \log \frac{1}{\tilde{V}_\rho(\xi)} + \int_0^{\kappa(\rho, \xi)} |\log(w)| e^{-w} dw \right] \end{aligned} \quad (132)$$

where the second inequality follows by lower-bounding $\log(\tilde{V}_\rho(\xi)w + \rho^{-1}) \geq \log(\tilde{V}_\rho(\xi)) + \log(w)$ and from the triangle inequality.

By using that the exponential function is nonnegative, by upper-bounding the integral by integrating to infinity, and by using $|\xi| > \xi_0$, we can further upper-bound (132), for sufficiently large ρ , by

$$J_4(\rho, \xi) \leq \frac{\tilde{V}_\rho(\xi)}{\tilde{V}_\rho(\xi) + \xi_0^2} \left[K + \log \frac{1}{\tilde{V}_\rho(\xi)} \right] \quad (133)$$

where we define

$$K \triangleq \int_0^\infty |\log(w)| e^{-w} dw = \gamma - 2 \text{Ei}(-1) \quad (134)$$

and where $\text{Ei}(\cdot)$ denotes the exponential integral function, i.e.,

$$\text{Ei}(-x) \triangleq - \int_x^\infty \frac{e^{-u}}{u} du. \quad (135)$$

Noting that the RHS of (133) is a continuous and bounded function of $0 < \tilde{V}_\rho(\xi) \leq 1$ that vanishes as $\tilde{V}_\rho(\xi)$ tends to zero, it follows from (113), (38a), and Lemma 10 that

$$\overline{\lim}_{\rho \rightarrow \infty} \mathbb{E}[J_4(\rho, \hat{H}_\rho) \mathbb{I}\{|\hat{H}_\rho| > \xi_0\}] \leq \overline{\lim}_{\rho \rightarrow \infty} \mathbb{E}[J_4(\rho, \hat{H}_\rho)] \leq 0 \quad (136)$$

thus proving (110b) for $i = 4$.

VIII. SUMMARY AND CONCLUSION

We have demonstrated that rate splitting can increase the well-known capacity lower bound (5) by Médard [2] of fading channels with imperfect channel-state information at the receiver. By computing the maximum of these bounds over all possible rate-splitting strategies, we have established a novel capacity lower bound which is larger than Médard's lower bound (5).

Viewing said capacity lower bound as a lower bound on the Gaussian-input mutual information, we have studied the high-SNR behavior of the novel bound under the assumption that the variance of the channel estimation error tends to zero with the SNR. We have shown that, for Gaussian fading, the rate-splitting bound is asymptotically tight in the sense that its difference to the Gaussian-input mutual information vanishes as the SNR tends to infinity. In contrast to Médard's lower bound, which is only asymptotically tight if the variance of the estimation error decays faster than the reciprocal of the SNR, the novel lower bound is asymptotically tight irrespective of the speed at which this variance decays.

In [5], Lapidoth and Shamai have shown that Médard's lower bound corresponds to the GMI for i.i.d. Gaussian codebooks and a scaled nearest neighbor decoder. From this mismatched decoding perspective, the combination of rate splitting (at the transmitter) and successive decoding (at the receiver) corresponds to a modification of the encoding/decoding rule which turns out advantageous in achieving higher transmission rates.

APPENDIX A

PROOF OF PROPOSITION 1

We expand the mutual information as

$$I(S; AS + B|A, C) = h(S|A, C) - h(S|AS + B, A, C). \quad (137)$$

Since, by assumption, S is zero-mean, variance- P , circularly-symmetric Gaussian and independent of (A, C) , the first entropy on the RHS of (137) is readily evaluated as

$$h(S|A, C) = h(S) = \log(\pi e P). \quad (138)$$

Conditioned on $(A, C) = (a, c)$, the second entropy can be upper-bounded as follows:

$$\begin{aligned} h(S|AS + B, A = a, C = c) &= h(S - \alpha_{A,C}(AS + B - \mu_{B|A,C}) \mid AS + B, A = a, C = c) \\ &\leq h(S - \alpha_{A,C}(AS + B - \mu_{B|A,C}) \mid A = a, C = c) \\ &\leq \log \left(\pi e \mathbb{E} \left[|S - \alpha_{A,C}(AS + B - \mu_{B|A,C})|^2 \mid A = a, C = c \right] \right) \end{aligned} \quad (139)$$

for any arbitrary $\alpha_{a,c} \in \mathbb{C}$, where $\mu_{B|a,c} \triangleq \mathbb{E}[B|A = a, C = c]$. Here the first inequality follows because conditioning cannot increase entropy, and the second inequality follows from the entropy-maximizing property of the Gaussian distribution. Combining (139) with (138) and (137) thus yields for every $(A, C) = (a, c)$ and $\alpha_{a,c}$

$$I(S; AS + B|A = a, C = c) \geq \log \frac{P}{\mathbb{E} \left[|S - \alpha_{A,C}(AS + B - \mu_{B|A,C})|^2 \mid A = a, C = c \right]}. \quad (140)$$

We choose $\alpha_{a,c}$ so that $\alpha_{a,c}(aS + B - \mu_{B|a,c})$ is the linear MMSE estimate of S , namely,

$$\alpha_{a,c} = \frac{\mathbb{E}[S(AS + B - \mu_{B|A,C})^* \mid A = a, C = c]}{\mathbb{E}[|AS + B - \mu_{B|A,C}|^2 \mid A = a, C = c]} = \frac{a^* P}{|a|^2 P + V_B(a, c)} \quad (141)$$

where $V_B(a, c)$ denotes the conditional variance of B conditioned on $(A, C) = (a, c)$. This yields

$$\mathbb{E} \left[|S - \alpha_{A,C}(AS + B - \mu_{B|A,C})|^2 \mid A = a, C = c \right] = P \frac{V_B(a, c)}{|a|^2 P + V_B(a, c)}. \quad (142)$$

Consequently, combining (142) with (140) gives

$$I(S; AS + B|A = a, C = c) \geq \log \left(1 + \frac{|a|^2 P}{V_B(a, c)} \right). \quad (143)$$

Proposition 1 follows then by averaging over (A, C) .

APPENDIX B
PROOF OF LEMMA 2

In order to prove Lemma 2, we shall demonstrate for every $L \in \mathbb{N}$ that, if the layerings $\mathbf{Q} \in \mathcal{Q}(P, L)$ and $\mathbf{Q}' \in \mathcal{Q}(P, L+1)$ satisfy

$$\{Q_1, \dots, Q_L\} \subset \{Q'_1, \dots, Q'_{L+1}\} \quad (144)$$

then $R[\mathbf{Q}] \leq R[\mathbf{Q}']$ with equality if, and only if, $\Pr\{\hat{H} \cdot \tilde{V}(\hat{H}) = 0\} = 1$. The general case where $\mathbf{Q}' \in \mathcal{Q}(P, L')$ for some arbitrary $L' > L$ follows then directly from the case $L' = L+1$ by applying the above result $(L' - L)$ times.

Let the element in \mathbf{Q}' that is not contained in \mathbf{Q} be at position $\tau = 1, \dots, L$, i.e.,⁷

$$Q_\ell = Q'_\ell, \quad \ell = 1, \dots, \tau - 1 \quad (145a)$$

and

$$Q_\ell = Q'_{\ell+1}, \quad \ell = \tau, \dots, L. \quad (145b)$$

We next express $\Gamma_{\ell, \mathbf{A}}(X^{\ell-1}, \hat{H})$ in (22) for some general layering \mathbf{A} as

$$\Gamma_{\ell, \mathbf{A}}(X^{\ell-1}, \hat{H}) = \frac{|\hat{H}|^2 (A_\ell - A_{\ell-1})}{\tilde{V}(\hat{H}) |\sum_{i < \ell} X_i|^2 + (|\hat{H}|^2 + \tilde{V}(\hat{H}))P - |\hat{H}|^2 A_\ell - \tilde{V}(\hat{H})A_{\ell-1} + N_0}. \quad (146)$$

Noting that for the layering \mathbf{Q} the term $|\sum_{i < \ell} X_i|^2$ has an exponential distribution with mean $Q_{\ell-1}$, whereas for the layering \mathbf{Q}' , it has an exponential distribution with mean $Q'_{\ell-1}$, and using (145a) and (145b), it can be easily verified that

$$\mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{Q}}(X^{\ell-1}, \hat{H})) \right] = \mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{Q}'}(X^{\ell-1}, \hat{H})) \right], \quad \ell = 1, \dots, \tau - 1 \quad (147)$$

and

$$\mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{Q}}(X^{\ell-1}, \hat{H})) \right] = \mathbb{E} \left[\log(1 + \Gamma_{\ell+1, \mathbf{Q}'}(X^\ell, \hat{H})) \right], \quad \ell = \tau + 1, \dots, L. \quad (148)$$

Subtracting $R[\mathbf{Q}]$ from $R[\mathbf{Q}']$ yields thus

$$\begin{aligned} R[\mathbf{Q}'] - R[\mathbf{Q}] &= \mathbb{E} \left[\log(1 + \Gamma_{\tau, \mathbf{Q}'}(X^{\tau-1}, \hat{H})) \right] \\ &\quad + \mathbb{E} \left[\log(1 + \Gamma_{\tau+1, \mathbf{Q}'}(X^\tau, \hat{H})) \right] - \mathbb{E} \left[\log(1 + \Gamma_{\tau, \mathbf{Q}}(X^{\tau-1}, \hat{H})) \right]. \end{aligned} \quad (149)$$

Since the random variables $X_1, \dots, X_\tau, \hat{H}$ are independent, we can express the second expectation as

$$\mathbb{E}_{X^{\tau-1}, \hat{H}} \left[\mathbb{E}_{X_\tau} \left[\log(1 + \Gamma_{\tau+1, \mathbf{Q}'}(X^\tau, \hat{H})) \right] \right] \quad (150)$$

where the subscript indicates over which random variables the expected value is computed. Using that, for every $\alpha > 0$, the function $x \mapsto \log(1 + \alpha/x)$ is strictly convex in $x \geq 0$, it follows from Jensen's inequality that, for every $X^{\tau-1} = x^{\tau-1}$ and $\hat{H} = \hat{h}$, the inner expectation is lower-bounded by

$$\mathbb{E}_{X_\tau} \left[\log(1 + \Gamma_{\tau+1, \mathbf{Q}'}(x^{\tau-1}, X_\tau, \hat{h})) \right] \geq \log(1 + \bar{\Gamma}_{\tau+1, \mathbf{Q}'}(x^{\tau-1}, \hat{h})) \quad (151)$$

⁷By the definition of a layering, we have $Q'_{L+1} = Q_L = P$, so the element in \mathbf{Q}' not contained in \mathbf{Q} cannot be at position $\tau = L+1$.

where we define

$$\bar{\Gamma}_{\tau+1, \mathbf{Q}'}(x^{\tau-1}, \hat{h}) \triangleq \frac{|\hat{h}|^2(Q'_{\tau+1} - Q'_\tau)}{\tilde{V} \left| \sum_{i < \tau} x_i \right|^2 + (|\hat{h}|^2 + \tilde{V})P - |\hat{h}|^2 Q'_{\tau+1} - \tilde{V} Q'_{\tau-1} + N_0} \quad (152)$$

and where the denominator of $\bar{\Gamma}_{\tau+1, \mathbf{Q}'}(x^{\tau-1}, \hat{h})$ is obtained by noting that X_τ has zero mean, so

$$\mathbb{E}_{X_\tau} \left[\left| \sum_{i < \tau} x_i + X_\tau \right|^2 \right] = \left| \sum_{i < \tau} x_i \right|^2 + Q'_\tau - Q'_{\tau-1}. \quad (153)$$

Since $\mathbf{Q}' \in \mathcal{Q}(P, L+1)$ implies that $\mathbb{E}[|X_\tau|^2] > 0$, the inequality in (151) is strict except in the trivial cases $\tilde{V}(\hat{h}) = 0$ or $\hat{h} = 0$. Combining (150) and (151) yields

$$\mathbb{E} \left[\log(1 + \Gamma_{\tau+1, \mathbf{Q}'}(X^\tau, \hat{H})) \right] \geq \mathbb{E} \left[\log(1 + \bar{\Gamma}_{\tau+1, \mathbf{Q}'}(X^{\tau-1}, \hat{H})) \right] \quad (154)$$

which together with (149) gives

$$R[\mathbf{Q}'] - R[\mathbf{Q}] \geq \mathbb{E} \left[\log(1 + \Gamma_{\tau, \mathbf{Q}'}(X^{\tau-1}, \hat{H})) \right] + \mathbb{E} \left[\log \left(\frac{1 + \bar{\Gamma}_{\tau+1, \mathbf{Q}'}(X^{\tau-1}, \hat{H})}{1 + \Gamma_{\tau, \mathbf{Q}}(X^{\tau-1}, \hat{H})} \right) \right] \quad (155)$$

with the inequality being strict except if $\Pr\{\hat{H} \cdot \tilde{V}(\hat{H}) = 0\} = 1$.

We next use (145a) and (145b) and the fact that $|\sum_{i < \tau} X_i|^2$ has an exponential distribution with mean $Q'_{\tau-1}$ under both layerings \mathbf{Q} and \mathbf{Q}' to evaluate the second expected value on the RHS of (155):

$$\begin{aligned} \mathbb{E} \left[\log \left(\frac{1 + \bar{\Gamma}_{\tau+1, \mathbf{Q}'}(X^{\tau-1}, \hat{H})}{1 + \Gamma_{\tau, \mathbf{Q}}(X^{\tau-1}, \hat{H})} \right) \right] &= \mathbb{E} \left[\log \left(\frac{T - |\hat{H}|^2 Q'_\tau - \tilde{V}(\hat{H}) Q'_{\tau-1} + N_0}{T - |\hat{H}|^2 Q'_{\tau-1} - \tilde{V}(\hat{H}) Q'_{\tau-1} + N_0} \right) \right] \\ &= -\mathbb{E} \left[\log \left(1 + \frac{|\hat{H}|^2 (Q'_\tau - Q'_{\tau-1})}{T - |\hat{H}|^2 Q'_{\tau-1} - \tilde{V}(\hat{H}) Q'_{\tau-1} + N_0} \right) \right] \end{aligned} \quad (156)$$

where we introduce

$$T \triangleq \tilde{V}(\hat{H}) \left| \sum_{i < \tau} X_i \right|^2 + (|\hat{H}|^2 + \tilde{V}(\hat{H}))P \quad (157)$$

for ease of exposition. Noting that

$$\frac{|\hat{H}|^2 (Q'_\tau - Q'_{\tau-1})}{T - |\hat{H}|^2 Q'_{\tau-1} - \tilde{V}(\hat{H}) Q'_{\tau-1} + N_0} = \Gamma_{\tau, \mathbf{Q}'}(X^{\tau-1}, \hat{H}) \quad (158)$$

yields that the RHS of (155) is zero, thus demonstrating that

$$R[\mathbf{Q}] \leq R[\mathbf{Q}'] \quad (159)$$

with equality if, and only if, $\Pr\{\hat{H} \cdot \tilde{V}(\hat{H}) = 0\} = 1$. This proves Lemma 2.

APPENDIX C PROOF OF LEMMA 7

We show that

$$\lim_{\mathbf{Q} \rightarrow \mathbf{Q}'} R[\mathbf{Q}] = R[\mathbf{Q}'] \quad (160)$$

where $\mathbf{Q} \rightarrow \mathbf{Q}'$ should be read as

$$\max_{\ell} |Q_\ell - Q'_\ell| \rightarrow 0. \quad (161)$$

To this end, we write $R[\mathbf{Q}]$ as

$$R[\mathbf{Q}] = \sum_{\ell=1}^L \mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{Q}}(W_{\ell}, \hat{H})) \right] \quad (162)$$

with

$$\Gamma_{\ell, \mathbf{Q}}(W_{\ell}, \hat{H}) = \frac{|\hat{H}|^2(Q_{\ell} - Q_{\ell-1})}{\tilde{V}(\hat{H})W_{\ell}Q_{\ell-1} + \tilde{V}(\hat{H})(Q_{\ell} - Q_{\ell-1}) + (|\hat{H}|^2 + \tilde{V}(\hat{H}))(P - Q_{\ell}) + N_0} \quad (163)$$

(assuming that $Q_0 = 0$) and

$$W_{\ell} \triangleq \begin{cases} 0, & \ell = 0 \\ \frac{1}{Q_{\ell-1}} \left| \sum_{i < \ell} X_i \right|^2, & \ell = 2, \dots, L. \end{cases} \quad (164)$$

Using that, with probability one,

$$0 \leq \log(1 + \Gamma_{\ell, \mathbf{Q}}) \leq \frac{|\hat{H}|^2 P}{N_0} \quad (165)$$

and that \hat{H} has finite variance, it follows from the Dominated Convergence Theorem [19] that

$$\begin{aligned} \lim_{\mathbf{Q} \rightarrow \mathbf{Q}'} \mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{Q}}(W_{\ell}, \hat{H})) \right] &= \mathbb{E} \left[\lim_{\mathbf{Q} \rightarrow \mathbf{Q}'} \log(1 + \Gamma_{\ell, \mathbf{Q}}(W_{\ell}, \hat{H})) \right] \\ &= \mathbb{E} \left[\log(1 + \Gamma_{\ell, \mathbf{Q}'}(W_{\ell}, \hat{H})) \right] \end{aligned} \quad (166)$$

where the last step follows by noting that $\mathbf{Q} \mapsto \log(1 + \Gamma_{\ell, \mathbf{Q}}(X^{\ell-1}, \hat{H}))$ is a continuous function of \mathbf{Q} . Combining (166) with (162) proves (160) and, hence, Lemma 7.

APPENDIX D

PROOF OF LEMMA 8

We first note that specializing Theorem 3 to the case $\hat{H} = \xi$ gives

$$R_M(\rho, \xi) \leq \mathbb{E}[R^*(\rho, W, \xi)], \quad (\rho > 0, x \in \mathbb{C}). \quad (167)$$

Consequently, we obtain from the definitions (95) and (94) of $\Sigma(\rho, \xi)$ and $\Delta(\rho, w, \xi)$ that

$$\begin{aligned} \Sigma(\rho, \xi) &\leq \mathbb{E}[\Delta(\rho, W, \xi)] \\ &= \log \left(\frac{\tilde{V}_{\rho}(\xi) + \rho^{-1}}{\tilde{\Phi}_{\rho}(\xi)} \right) - \mathbb{E} \left[\log \left(W + \frac{1}{\rho \tilde{\Phi}_{\rho}(\xi)} \right) \right]. \end{aligned} \quad (168)$$

The second term on the RHS of (168) gives [20]

$$\mathbb{E} \left[\log \left(W + \frac{1}{\rho \tilde{\Phi}_{\rho}(\xi)} \right) \right] = \log \left(\frac{1}{\rho \tilde{\Phi}_{\rho}(\xi)} \right) - e^{\frac{1}{\rho \tilde{\Phi}_{\rho}(\xi)}} \text{Ei} \left(-\frac{1}{\rho \tilde{\Phi}_{\rho}(\xi)} \right). \quad (169)$$

Consequently, we have

$$\begin{aligned}
\mathbb{E}[\Delta(\rho, W, \xi)] &= \log(1 + \rho \tilde{V}_\rho(\xi)) + e^{\frac{1}{\rho \tilde{\Phi}_\rho(\xi)}} \text{Ei}\left(-\frac{1}{\rho \tilde{\Phi}_\rho(\xi)}\right) \\
&= \log\left(1 + \rho \tilde{\Phi}_\rho(\xi) \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)}\right) + e^{\frac{1}{\rho \tilde{\Phi}_\rho(\xi)}} \text{Ei}\left(-\frac{1}{\rho \tilde{\Phi}_\rho(\xi)}\right) \\
&\leq \log\left(1 + \rho \tilde{\Phi}_\rho(\xi) \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)}\right) + \text{Ei}\left(-\frac{1}{\rho \tilde{\Phi}_\rho(\xi)}\right) \\
&\triangleq g\left(\rho \tilde{\Phi}_\rho(\xi); \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)}\right)
\end{aligned} \tag{170}$$

where the inequality follows because $\text{Ei}(-x)$ is negative for $x > 0$ and $e^x \geq 1$, $x > 0$; and where the last step should be viewed as the definition of $g(\cdot; \cdot)$.

For a fixed a , the function $t \mapsto g(t; a)$ satisfies [4, Section VI-B]

$$\lim_{t \rightarrow \infty} g(t; a) = \gamma + \log(a). \tag{171}$$

We next show that, for every $a \geq 1$, the function $t \mapsto g(t; a)$ is monotonically increasing. Indeed, we have

$$\begin{aligned}
\frac{\partial}{\partial t} g(t; a) &= \frac{e^{-\frac{1}{t}}}{(1+at)t} \left[e^{\frac{1}{t}} at - 1 - at \right] \\
&\geq \frac{e^{-\frac{1}{t}}}{(1+at)t} [a - 1] \\
&\geq 0, \quad a \geq 1
\end{aligned} \tag{172}$$

where the second step follows from the lower bound $e^{\frac{1}{t}} \geq 1 + \frac{1}{t}$, $t \geq 0$.

Due to the entropy-maximizing property of Gaussian random variables, we have $\tilde{V}_\rho(\xi) \geq \tilde{\Phi}_\rho(\xi)$, which implies that $\tilde{V}_\rho(\xi)/\tilde{\Phi}_\rho(\xi) \geq 1$. It thus follows from (170)–(172) that

$$\begin{aligned}
\Sigma(\rho, \xi) &\leq g\left(\rho \tilde{\Phi}_\rho(\xi); \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)}\right) \\
&\leq \lim_{t \rightarrow \infty} g\left(t; \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)}\right) \\
&= \gamma + \log\left(\frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)}\right), \quad (\rho > 0, \xi \in \mathbb{C})
\end{aligned} \tag{173}$$

Maximizing the RHS of (173) over $\xi \in \mathbb{C}$ and computing the limit as ρ tends to infinity gives

$$\begin{aligned}
\overline{\lim}_{\rho \rightarrow \infty} \sup_{\xi \in \mathbb{C}} \Sigma(\rho, \xi) &\leq \overline{\lim}_{\rho \rightarrow \infty} \log\left(\sup_{\xi \in \mathbb{C}} \frac{\tilde{V}_\rho(\xi)}{\tilde{\Phi}_\rho(\xi)}\right) + \gamma \\
&\leq \gamma + \log(M)
\end{aligned} \tag{174}$$

where the last step follows from the continuity of $x \mapsto \log(x)$ and from (38c). This proves Lemma 8.

APPENDIX E
PROOF OF LEMMA 9

For any $\epsilon > 0$ we have

$$\Pr\{|H| > \xi_0 + \epsilon\} = \Pr\{|H| > \xi_0 + \epsilon, |\hat{H}_\rho| \leq \xi_0\} + \Pr\{|H| > \xi_0 + \epsilon, |\hat{H}_\rho| > \xi_0\} \quad (175)$$

$$\leq \Pr\{|H - \hat{H}_\rho| > \epsilon, |\hat{H}_\rho| \leq \xi_0\} + \Pr\{|\hat{H}_\rho| > \xi_0\} \quad (176)$$

where the inequality follows by upper-bounding the first probability on the RHS of (175) using the triangle inequality and by upper-bounding the second probability on the RHS of (175) using the law of total probability. Using the law of total probability and Markov's inequality [21], the first term on the RHS of (176) can be further upper-bounded by

$$\begin{aligned} \Pr\{|H - \hat{H}_\rho| > \epsilon, |\hat{H}_\rho| \leq \xi_0\} &\leq \Pr\{|H - \hat{H}_\rho| > \epsilon\} \\ &\leq \frac{\mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)]}{\epsilon^2}. \end{aligned} \quad (177)$$

Combining (177) with (176) gives

$$\Pr\{|H| > \xi_0 + \epsilon\} \leq \frac{\mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)]}{\epsilon^2} + \Pr\{|\hat{H}_\rho| > \xi_0\}. \quad (178)$$

Since, by the lemma's assumptions, we have $\lim_{\rho \rightarrow \infty} \mathbb{E}[\tilde{V}_\rho(\hat{H}_\rho)] = 0$, taking the limit inferior for $\rho \rightarrow \infty$ on either side of (178) yields

$$\Pr\{|H| > \xi_0 + \epsilon\} \leq \lim_{\rho \rightarrow \infty} \Pr\{|\hat{H}_\rho| > \xi_0\}. \quad (179)$$

We next note that, for sufficiently large ρ , the conditional distribution of \tilde{H}_ρ , conditioned on $\hat{H}_\rho = \xi$, must be absolutely continuous with respect to the Lebesgue measure, since otherwise $\tilde{\Phi}_\rho(\xi) = 0$, contradicting the Lemma's assumption that $\tilde{\Phi}_\rho(\xi)$ satisfies (38c). This implies that the cumulative distribution function of $|H| = |\hat{H}_\rho + \tilde{H}_\rho|$ is continuous. Consequently, we have

$$\lim_{\epsilon \downarrow 0} \Pr\{|H| > \xi_0 + \epsilon\} = \Pr\{|H| > \xi_0\} \quad (180)$$

which together with (179) gives

$$\lim_{\rho \rightarrow \infty} \Pr\{|\hat{H}_\rho| > \xi_0\} \geq \Pr\{|H| > \xi_0\} \quad (181)$$

upon letting ϵ tend to zero from above. Using the continuity of the cumulative distribution function of $|H|$, Lemma 9 follows from (181) by noting that

$$\lim_{\xi_0 \downarrow 0} \Pr\{|H| > \xi_0\} = \Pr\{|H| > 0\} = 1. \quad (182)$$

APPENDIX F

PROOF OF LEMMA 10

For every family of random variables \mathcal{Y}_ρ parametrized by $\rho > 0$ and taking values on $(0, 1]$, we have by Markov's inequality

$$\Pr \{ \mathcal{Y}_\rho > \nu \} \leq \frac{\mathbb{E}[\mathcal{Y}_\rho]}{\nu}, \quad \text{for every } \nu > 0. \quad (183)$$

Using that $\lim_{\rho \rightarrow \infty} \mathbb{E}[\mathcal{Y}_\rho] = 0$, we thus have

$$\lim_{\rho \rightarrow \infty} \Pr \{ \mathcal{Y}_\rho > \nu \} = 0, \quad \text{for every } \nu > 0 \quad (184)$$

or equivalently, $\lim_{\rho \rightarrow \infty} \Pr \{ \mathcal{Y}_\rho \leq \nu \} = 1$. We upper-bound $\mathbb{E}[f(\mathcal{Y}_\rho)]$ for any $\nu > 0$ as

$$\begin{aligned} \mathbb{E}[f(\mathcal{Y}_\rho)] &= \mathbb{E}[f(\mathcal{Y}_\rho) \mathbf{I}\{\mathcal{Y}_\rho \leq \nu\}] + \mathbb{E}[f(\mathcal{Y}_\rho) \mathbf{I}\{\mathcal{Y}_\rho > \nu\}] \\ &\leq \sup_{0 < t \leq \nu} f(t) \Pr\{\mathcal{Y}_\rho \leq \nu\} + \sup_{\nu < t \leq 1} f(t) \Pr\{\mathcal{Y}_\rho > \nu\} \end{aligned} \quad (185a)$$

Similarly, we lower-bound $\mathbb{E}[f(\mathcal{Y}_\rho)]$ for any $\nu > 0$ as

$$\mathbb{E}[f(\mathcal{Y}_\rho)] \geq \inf_{0 < t \leq \nu} f(t) \Pr\{\mathcal{Y}_\rho \leq \nu\} + \inf_{\nu < t \leq 1} f(t) \Pr\{\mathcal{Y}_\rho > \nu\} \quad (185b)$$

Since $f(\cdot)$ is bounded, and due to (184), taking limits for $\rho \rightarrow \infty$ in (185a) and (185b) gives

$$\inf_{0 < t \leq \nu} f(t) \leq \lim_{\rho \rightarrow \infty} \mathbb{E}[f(\mathcal{Y}_\rho)] \leq \overline{\lim}_{\rho \rightarrow \infty} \mathbb{E}[f(\mathcal{Y}_\rho)] \leq \sup_{0 < t \leq \nu} f(t). \quad (186)$$

Taking the limit as ν tends to zero from above, we finally obtain

$$\lim_{\rho \rightarrow \infty} \mathbb{E}[f(\mathcal{Y}_\rho)] = \lim_{t \downarrow 0} f(t) = f_0 \quad (187)$$

which proves Lemma 10.

ACKNOWLEDGMENT

Fruitful discussions with Amos Lapidoth are gratefully acknowledged.

REFERENCES

- [1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, 1998.
- [2] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. John Wiley & Sons, 1991.
- [4] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [5] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?" *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [6] A. Dembo and T. Cover, "Information theoretic inequalities," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.
- [7] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

- [8] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
- [9] J. Doob, *Stochastic Processes*. John Wiley & Sons, 1990.
- [10] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [11] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat-fading channels," *IEEE Transactions on Signal Processing*, vol. 21, no. 5, pp. 1403–1418, May 2004.
- [12] A. Lozano, "Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5020–5029, Dec. 2008.
- [13] A. Asyhari, T. Koch, and A. Guillén i Fàbregas, "Nearest neighbour decoding and pilot-aided channel estimation in stationary Gaussian flat-fading channels," in *Proc. IEEE International Symposium on Information Theory*, August 2011, pp. 2786–2790.
- [14] S. Ohno and G. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 712–720, Oct. 2002.
- [15] G. Kaplan and S. Shamai (Shitz), "Information rates of compound channels with application to antipodal signaling in a fading environment," *AEU*, vol. 47, no. 4, pp. 228–239, 1993.
- [16] I. G. Stiglitz, "Coding for a class of unknown channels," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 189–195, Apr. 1966.
- [17] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and wide-band limit," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.
- [18] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1439–1452, Sept. 1996.
- [19] W. Rudin, *Real and complex analysis*, 3rd ed. New York: McGraw-Hill, 1987.
- [20] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Elsevier/Academic Press, Amsterdam, Jan. 2007.
- [21] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.